The Annals of Probability
2011, Vol. 39, No. 3, 985–1026
DOI: 10.1214/10-AOP575

DISCREPANCY, CHAINING AND SUBGAUSSIAN PROCESSES

By Shahar Mendelson

Israel Institute of Technology and The Australian National University

We show that for a typical coordinate projection of a subgaussian class of functions, the infimum over signs $\inf_{(\varepsilon_i)} \sup_{f\in F}|\sum_{i=1}^k \varepsilon_i f(X_i)|$ is asymptotically smaller than the expectation over signs as a function of the dimension k, if the canonical Gaussian process indexed by F is continuous. To that end, we establish a bound on the discrepancy of an arbitrary subset of $\mathbb{R}^k$ using properties of the canonical Gaussian process the set indexes, and then obtain quantitative structural information on a typical coordinate projection of a subgaussian class.

1. Introduction. The geometric structure of a random coordinate projection of a class of functions plays an important role in Empirical Processes theory, where it is used to determine whether the uniform law of large numbers or the uniform central limit theorem is satisfied by the given class. Indeed, if F is a class of functions on a probability space (Ω,μ), and if $\sigma=(X_1,\dots,X_k)$ is an independent sample distributed according to μ^k, then the "complexity" of the random set

$$P_\sigma F=\{(f(X_1),\dots,f(X_k)) : f\in F\}\subset\mathbb{R}^k$$

is the key parameter in addressing both these questions. In this context, if $(\varepsilon_i)_{i=1}^k$ are independent, symmetric, $\{-1,1\}$-valued random variables, then the complexity is governed by the expectation of the supremum of the Bernoulli process indexed by $P_\sigma F$, defined by

$$\mathbb{E}_\varepsilon \sup_{f\in F}\left|\sum_{i=1}^k \varepsilon_i f(X_i)\right| = \mathbb{E}_\varepsilon \sup_{v\in P_\sigma F}\left|\sum_{i=1}^k \varepsilon_i v_i\right|, \tag{1.1}$$

and in particular, on the way this expectation grows as a function of k for a typical sample of cardinality k (see, e.g., [3, 8, 19] and references therein).

Received October 2008; revised June 2010.
AMS 2000 subject classifications. 60C05, 60G15, 60D05.
Key words and phrases. Discrepancy, generic chaining.

The structure of such coordinate projections is central to questions in Asymptotic Geometric Analysis as well. For example, let $K \subset \mathbb{R}^d$ be a convex, symmetric set (i.e., if $x \in K$ then $-x \in K$) and put $F = \{\langle x, \cdot\rangle : x \in K\}$ to be the class of linear functionals indexed by K. If μ is a measure on $\mathbb{R}^d$, then $P_\sigma F$ is the set ΓK, where Γ is the random operator $\Gamma = \sum_{i=1}^k \langle X_i, \cdot\rangle e_i$. Fundamental questions on the geometry of convex, symmetric sets, such as Dvoretzky's theorem and low-M^* estimates have been answered by obtaining accurate, quantitative information on the structure of such coordinate projections, and by using very similar complexity parameters to (1.1) (e.g., [8, 15]).

For both these reasons, a lot of effort has been invested in understanding various notions of complexity for a typical coordinate projection of a class of functions. A well studied direction is to obtain quantitative estimates on the way in which (1.1) is related to two other complexity parameters, the combinatorial dimension and covering numbers.

Roughly speaking, the combinatorial dimension of $T \subset \mathbb{R}^k$ at scale ε, denoted by $\mathrm{VC}(T, \varepsilon)$, is the largest dimension of a coordinate projection of T that contains a "cube" of scale ε (see Definition 6.2 for an exact formulation). If (T, d) is a metric space then the covering number at scale ε, which we denote by $N(\varepsilon, T, d)$, is the smallest cardinality of a subset $\{y_1, \ldots, y_m\} \subset T$ such that for every $t \in T$, there is some y_i for which $d(t, y_i) < \varepsilon$.

Connections between (1.1) and the combinatorial dimension or the covering numbers of $P_\sigma F$ are rather well understood. For example, a straightforward chaining argument (see, e.g., [19]) shows that for every $T \subset \mathbb{R}^k$,

$$\mathbb{E}_\varepsilon \sup_{t \in T} \left| \sum_{i=1}^k \varepsilon_i t_i \right| \leq c \int_0^{\mathrm{diam}(T)} \sqrt{\log N(\varepsilon, T, \ell_2^k)}\, d\varepsilon, \tag{1.2}$$

where ℓ_2^k is the Euclidean metric on $\mathbb{R}^k$, $\mathrm{diam}(T)$ is the diameter with respect to the same metric and c is an absolute constant, independent of the dimension k and of the set T. Some of the other relations between these parameters are far more involved. First, controlling the L_2 covering numbers using the combinatorial dimension was resolved in [12], where it was shown that if T is a subset of the unit cube B_∞^k and μ is any probability measure on $\{1, \ldots, k\}$, then for every $0 < \varepsilon < 1$,

$$N(\varepsilon, T, L_2(\mu)) \leq \left(\frac{5}{\varepsilon}\right)^{K \cdot \mathrm{VC}(T, c\varepsilon)},$$

where K and c are absolute constants. Also, the solution of the sign embedding of ℓ_1^k problem (see [12]) implies that if $T \subset B_\infty^k$ and $\mathbb{E}\sup_{t \in T} |\sum_{i=1}^k \varepsilon_i t_i| \geq \delta k$, then $\mathrm{VC}(T, c_1\delta) \geq c_2 \delta^2 k$. In other words, under a normalization condition ($T \subset B_\infty^k$), the only reason that $\mathbb{E}\sup_{t \in T} |\sum_{i=1}^k \varepsilon_i t_i|$ is almost extremal is that T contains a high-dimensional cubic structure.

In this article, we study a related geometric parameter—the discrepancy of a typical coordinate projection. Discrepancy was introduced in a combinatorial context (see the book [11] for an extensive survey on this topic) and is defined as follows.

DEFINITION 1.1. If $T \subset \mathbb{R}^k$, then the discrepancy of T is

$$\mathrm{disc}(T) = \inf_{(\varepsilon_i)_{i=1}^k} \sup_{t \in T} \left| \sum_{i=1}^k \varepsilon_i t_i \right|,$$

and the infimum is taken with respect to all signs $(\varepsilon_i)_{i=1}^k \in \{-1,1\}^k$.

We denote by $\mathrm{Hdisc}(T)$ the hereditary discrepancy of T, given by

$$\sup_{I \subset \{1,\ldots,k\}} \mathrm{disc}(P_I T),$$

where $P_I T = \{(t_i)_{i \in I} : t \in T\}$ is the coordinate projection of T onto I.

Observe that if $\mathrm{absconv}(T)$ is the convex hull of $T \cup -T$, then $\mathrm{disc}(T) = \mathrm{disc}(\mathrm{absconv}(T))$. Hence, from the geometric viewpoint, the discrepancy of T is proportional with a constant $\sqrt{k}$ to the minimal width of $\mathrm{absconv}(T)$ in a direction of a vertex of the combinatorial cube $\{-1,1\}^k$. The hereditary discrepancy is governed by a similar minimal width, but of the "worst" coordinate projection of $\mathrm{absconv}(T)$.

Our goal here is to study the discrepancy using the covering numbers and the combinatorial dimension of T, but we will focus on sets T that are random coordinate projections of a class of function F, which gives them more structure. A natural question in this context is to identify conditions on F under which there is a gap between $\mathrm{disc}(P_\sigma F)$ and $\mathbb{E}_\varepsilon \sup_{v \in P_\sigma F} |\sum_{i=1}^k \varepsilon_i v_i|$ for a typical coordinate projection of F, as a function of the sample size k. To that end, we will develop dimension dependent bounds on the discrepancy of $P_\sigma F$ (and in particular, bounds that are not asymptotic).

Note that the metric structure of $T \subset \mathbb{R}^k$ is not enough to determine its discrepancy. Indeed, if $e_1 = (1,0,\ldots,0) \in \mathbb{R}^k$ and $T_1 = \{0, e_1\}$ then $\mathrm{disc}(T_1) = 1$. On the other hand, if $T_2 = \{0, \sum_{i=1}^k e_i/\sqrt{k}\}$, which is linearly isometric to T_1, then $\mathrm{disc}(T_2) \le 1/\sqrt{k}$. The reason for the large gap in the discrepancy between the two isometric sets is that T_2 consists of a vector that is "well spread" while T_1 consists of a "peaky" vector with respect to the underlying coordinate structure. In that sense, T_2 is in a much better position than T_1. Note that in this example, $\mathbb{E}\sup_{t \in T_1} |\sum_{i=1}^k \varepsilon_i t_i| = \mathbb{E}\sup_{t \in T_2} |\sum_{i=1}^k \varepsilon_i t_i| = 1$—and for the set T_2, which is in a "good position" there is gap between the expectation for signs and the discrepancy.

We will show that this is a general phenomenon: it is well known that $\mathbb{E}_\varepsilon \sup_{t \in T} |\sum_{i=1}^k \varepsilon_i t_i|$ is determined by the Euclidean metric structure of T

(up to a logarithmic factor in the dimension), and therefore, it is almost invariant under a linear isometry (i.e., a change in the coordinate structure). Thus, the expectation almost does not change when applying an isometry or a good isomorphism of ℓ_2^k. As we will explain here, the situation with the discrepancy is rather different and the position of the set matters a great deal. Since the sets T that we will be interested in are not arbitrary but have some structure—as random coordinate projections of well behaved function classes, they will be much closer in nature to T_2 than to T_1.

Our main result is that if the canonical Gaussian process indexed by $F \subset L_2(\mu)$ is continuous and if the class satisfies a subgaussian condition [i.e., if the $\psi_2(\mu)$ norm is equivalent to the $L_2(\mu)$ norm on F, see Definition 2.4], then a typical coordinate projection of F behaves as a set of vectors in a "general position." As such, and just like the set T_2, a typical coordinate projection exhibits certain shrinking properties that will be explained in Section 4, and which causes the discrepancy of such a set to be much smaller than the average over signs.

THEOREM A. *Let $F \subset L_2(\mu)$ be a class of mean zero functions. Assume further that the canonical Gaussian process indexed by F is continuous and that the $L_2(\mu)$ and $\psi_2(\mu)$ norms are equivalent on F. Then* $\mathrm{Hdisc}(P_\sigma F)/\sqrt{k} \to 0$ *in probability.*

To put Theorem A in the right perspective, observe that if the ψ_2 and L_2 norms are equivalent on a class of mean zero functions F, then for every integer k there is a subset of Ω^k of probability at least c on which

$$\mathbb{E}_\varepsilon \sup_{f\in F}\left|\sum_{i=1}^k \varepsilon_i f(X_i)\right| \geq c_1\sqrt{k}\sigma_F, \tag{1.3}$$

where c depends only on the equivalence constant between the ψ_2 and L_2 norms on F, c_1 is an absolute constant and $\sigma_F = \sup_{f\in F}(\mathbb{E}f^2)^{1/2}$. Hence, there is a true gap between the discrepancy and the mean of a typical coordinate projection.

Although the formulation of Theorem A is asymptotic, the result itself is quantitative in nature, as a function of the dimension of the coordinate projection. The proof of Theorem A is, in fact, a dimension dependent estimate on the sequences $(\alpha_{k,\delta})_{k=1}^\infty$, for which, with probability at least $1-\delta$, $\mathrm{Hdisc}(P_\sigma F) \leq \sqrt{k}a_{k,\delta}$. We will show that the sequences $(\alpha_{k,\delta})_{k=1}^\infty$ are given using metric parameters that measure the continuity of the Gaussian process indexed by F—Talagrand's $\gamma_{2,s}$ functionals [18]. The $\gamma_{2,s}$ functionals will be defined in Section 2, but for now let us mention that under mild measurability assumptions on the class, the canonical Gaussian process indexed by F is continuous if and only if $\lim_{s\to\infty}\gamma_{2,s}(F, L_2(\mu)) = 0$.

We will prove that for every $0 < \rho < 1/2$ and $0 < \delta < 1$ there are constants c and C that depend on ρ, δ and on the equivalence constant between the $\psi_2(\mu)$ and $L_2(\mu)$ norms on F, such that for every k,

$$\alpha_{k,\delta} \leq C \sup_{1 \leq n \leq k} \sqrt{\frac{n}{k}} \left(\gamma_{2,\log_2 \log_2 cn}(F, L_2(\mu)) \cdot \sqrt{\log(ek/n)} + D \frac{\log k}{n^{1/2-\rho}} \right), \tag{1.4}$$

where $D = \mathrm{diam}(F, L_2(\mu))$ is the diameter of F with respect to the $L_2(\mu)$ norm. And, in particular, under the assumptions of Theorem A, for every $0 < \delta < 1$, $\lim_{k \to \infty} \alpha_{k,\delta} = 0$. Moreover, the proof of Theorem A actually shows that for every k, with μ^k-probability of at least $1 - \delta$,

$$\mathrm{disc}(P_\sigma F) \leq C(\sqrt{k}\gamma_{2,\log_2 \log_2 ck}(F, L_2(\mu)) + k^\rho D),$$

where C and c depend on ρ, δ and the equivalence constant between the $L_2(\mu)$ and $\psi_2(\mu)$ norms on F.

The proof of Theorem A is based on two ingredients. The first is a new estimate on the discrepancy of an arbitrary set $T \subset \mathbb{R}^k$. It is a combination of the entropy method, which is often used to control the combinatorial discrepancy (see, e.g., [1, 11, 16]), and Talagrand's generic chaining mechanism [18], which was introduced to establish the connection between the $\gamma_{2,s}$ functionals and the continuity of Gaussian processes. The combination of these two methods will be explained in Section 3. It allows one to find a good choice of signs for roughly $k/2$ coordinates, while the error incurred by considering the sum taken only on these coordinates is determined by the $\gamma_{2,s}$ functional for $s \sim \log_2 \log_2 k$. Repeating this argument, one obtains a bound on the discrepancy of T in terms of a sum of $\gamma_{2,s}$ functionals of coordinate projections of T and for values s that depend on the dimension of each projection, and those dimensions decrease quickly.

The second component required for the proof of Theorem A is that the sets we are interested in are not general. We will obtain a structural result on a typical $P_\sigma F$ that allows us to bound the $\gamma_{2,s}$ functionals of its coordinate projections using the $L_2(\mu)$ structure of F.

Indeed, we will show that if the $L_2(\mu)$ and $\psi_2(\mu)$ norms are equivalent on F then a typical coordinate projection $P_\sigma F$ has a rather regular structure—it is a subset of a Minkowski sum of two sets. The first one is small, with a bounded diameter in the weak ℓ_2 space $\ell_{2,\infty}^k$. The other set is a subset of $P_\sigma F$ itself and can be viewed as a set of vectors in a "general position." In particular, further coordinate projections of the latter set shrink distances between any two of its elements.

The structural result we obtain is of independent interest and can be used to derive information on the geometry of convex sets. For example, consider a symmetric probability measure μ on $\mathbb{R}^n$. We say that μ isotropic and L-subgaussian if a random vector X distributed according to μ satisfies that

for every $x \in \mathbb{R}^n$,

$$\mathbb{E}|\langle X, x\rangle|^2 = |x|^2 \quad \text{and} \quad \|\langle X, x\rangle\|_{\psi_2} \le L|x|.$$

Simple examples of isotropic, L-subgaussian measures on $\mathbb{R}^n$ are the Gaussian measure and the uniform measure on the vertices of the cube $\{-1,1\}^n$, where in both cases L can be taken to be an absolute constant, independent of the dimension.

Let $(X_i)_{i=1}^k$ be independent random vectors, distributed according to μ and consider the random operator $\Gamma : \mathbb{R}^n \to \mathbb{R}^k$ defined by $\Gamma = \sum_{i=1}^k \langle X_i, \cdot\rangle e_i$.

Corollary B. *For any $L > 0$ there are constants c_1, c_2 and c_3 that depend only on L, for which the following holds. Let $T \subset \mathbb{R}^n$ and set $V = k^{-1/2}\Gamma T$. Then, for every $u > c_1$, with probability at least $1 - 2\exp(-c_2 u)$, for every $I \subset \{1, \ldots, k\}$,*

$$\mathbb{E}_g \sup_{v \in V} \left|\sum_{i \in I} g_i v_i\right| \le c_3 u \sqrt{\frac{|I|}{k} \log\left(\frac{ek}{|I|}\right)}\, \mathbb{E}_g \sup_{t \in T} \left|\sum_{i=1}^n g_i t_i\right|,$$

where (g_i) are independent, standard Gaussian random variables, and both expectations are taken with respect to those variables.

Corollary B shows that the random operator Γ maps an arbitrary T to a set of vectors in a "general position" in a strong sense, since it implies that for most vectors in V, mutual distances are shrunk by any further coordinate projection. Let us note that we will prove a stronger result than Corollary B, namely that the $\gamma_{2,s}$ functionals associated with V display this type of shrinking phenomenon.

The final result we present has to do with the reverse direction of Theorem A. Assume that $H \subset L_2(\mu)$ is a convex, symmetric set, which satisfies that the canonical Gaussian process $\{G_h : h \in H\}$ is bounded and that the $L_2(\mu)$ and $\psi_2(\mu)$ norms are equivalent on H. We will show that if the logarithm of the $L_2(\mu)$ covering numbers of H grows like $1/\varepsilon^2$ then for a typical sample $\sigma = (X_1, \ldots, X_k)$ selected according to μ^k,

$$\mathrm{VC}(P_\sigma H, c_1/\sqrt{k}) \ge c_2 k.$$

It is standard to verify (see Lemma 6.5) that if $T \subset \mathbb{R}^k$, then

$$\mathrm{Hdisc}(T) \ge \sup_{\delta > 0} \delta\, \mathrm{VC}(\mathrm{absconv}(T), \delta).$$

Therefore, if F is a class of mean-zero functions and $H = \mathrm{absconv}(F)$ satisfies the above, then $\mathrm{Hdisc}(P_\sigma F) \ge c\sqrt{k}$, complementing the upper bound established in Theorem A.

Although this is not exactly the reverse direction of Theorem A, it is very close to it. Indeed, if $F \subset L_2(\mu)$ indexes a bounded Gaussian process then so does $H = \mathrm{absconv}(F)$, and the logarithm of the covering numbers of H cannot grow faster than $O(1/\varepsilon^2)$. On the other hand, if the log-covering numbers grow a little slower, even by a suitable logarithmic factor, then $\gamma_{2,s}(F, L_2(\mu)) \to 0$. In fact, this is as close as one can get to a covering numbers characterization of the fact that $\gamma_{2,s}(F, L_2) \to 0$ (see, e.g., [3]).

This result not only shows that $\mathrm{Hdisc}(P_\sigma F)$ is large if the Gaussian process F indexes is bounded but not continuous, it also shows why. Under a boundedness assumption on the Gaussian process [which implies that $\mathrm{Hdisc}(P_\sigma F)/\sqrt{k}$ is bounded], the reason the hereditary discrepancy of $P_\sigma F$ is extremal is because a typical coordinate projection of $\mathrm{absconv}(F)$ contains a high dimensional, large cubic structure, and that forces the hereditary discrepancy to be large. The proof of this result, which is presented in Section 6, is based on the observation that if F is convex and symmetric then the richness of F at scale $\sim 1/\sqrt{k}$ is exhibited by the existence of cubes of scale $\sim 1/\sqrt{k}$ and of dimension $\sim k$ in a typical coordinate projection of F of dimension k. It thus should be viewed as a "small scale" version of the Sign Embedding theorem which was mentioned above.

Unfortunately, the optimal estimate in the Sign Embedding theorem cannot be used directly in our case, firstly because $P_\sigma F$ is unlikely to be a subset of B_∞^k, and secondly, because a typical coordinate projection of F satisfies that

$$\mathbb{E} \sup_{v \in P_\sigma F} \left| \sum_{i=1}^k \varepsilon_i v_i \right| \sim \sqrt{k}.$$

Hence, the optimal estimate in the Sign Embedding theorem has to be used for $\delta \sim 1/\sqrt{k}$, and that only ensures that $P_\sigma F$ contains a cube of scale $\sim 1/\sqrt{k}$ and of constant dimension, which is far from what we need.

The proof of the existence of a cube in $P_\sigma F$ is based on two localization arguments, one with respect to the L_2 norm and the other with respect to the L_∞ norm. The first localization shows that if the $L_2(\mu)$ covering number of F at scale $\sim 1/\sqrt{k}$ is of the order of $\exp(c_1 k)$ then the richness of a typical coordinate projection of F of dimension $\sim k$ originates from the set

$$F_1 = F \cap \frac{c_2}{\sqrt{k}} B(L_2(\mu)), \tag{1.5}$$

that is, functions in F of $L_2(\mu)$ norm at most $O(1/\sqrt{k})$. In the second localization, one shows that the complexity of a typical coordinate projection actually comes from a further pointwise truncation of the functions in F, and $B(L_2(\mu))$ in (1.5) can essentially be replaced by $B(L_\infty(\mu))$—the unit ball in $L_\infty(\mu)$.

This article is organized as follows. In Section 2, we present further preliminaries, most of them concerning subgaussian variables and the $\gamma_{2,s}$ functionals. In Section 3, we develop bounds on the discrepancy of an arbitrary subset of $\mathbb{R}^n$. Section 4 is devoted to the proof of the structural results on coordinate projections of subgaussian processes and its corollaries, including Corollary B. Theorem A is proved in Section 5 and its converse and the resulting lower bound on the hereditary discrepancy of a typical coordinate projection is proved in Section 6.

2. Preliminaries. Throughout, absolute constants (i.e., fixed, positive numbers) will be denoted by C, c, c_1 etc. Their values may change from line to line. We denote by $C(a), c(a)$ constants that depend only on the parameter a and we set $\kappa_1, \kappa_2, \dots$ to be constants that will remain fixed throughout the article. By $a \sim b$, we mean that there are constants c and C such that $ca \le b \le Ca$, and we write $b \lesssim a$ if $b \le Ca$.

We will consider a single, fixed Euclidean structure on all finite-dimensional spaces $\mathbb{R}^n$ and denote the corresponding Euclidean norms by $|\cdot|$ without specifying the dimension. With a minor abuse of notation, the cardinality of a set and the absolute value are denoted in the same way.

If E is a normed space, let $B(E)$ be its unit ball, and for $E = \ell_p^n$ we denote the unit ball by B_p^n. If $\sigma = (X_1, \dots, X_k) \in \Omega^k$ let $\mu_k = k^{-1} \sum_{i=1}^k \delta_{X_i}$ be the empirical measure supported on σ, set L_2^k to be the corresponding L_2 space, and for $I \subset \{1, \dots, k\}$ let ℓ_2^I be the coordinate subspace of ℓ_2^k spanned by $(e_i)_{i \in I}$.

The situation we will study here is as follows. Let F be a class of real valued functions on a probability space (Ω, μ), let $X_1, \dots, X_k$ be independent random variables distributed according to μ and set $\sigma = (X_1, \dots, X_k)$. Let $P_\sigma F = \{(f(X_i))_{i=1}^k : f \in F\} \subset \mathbb{R}^k$ be the coordinate projection of F defined by σ and for every $I \subset \{1, \dots, k\}$ let $P_I^\sigma F = \{(f(X_i))_{i \in I} : f \in F\} \subset \mathbb{R}^{|I|}$ be the coordinate projection of F on the subset of coordinates $(X_i)_{i \in I}$. Sometimes, for the sake of simplicity, we shall omit the superscript σ.

2.1. *Subgaussian processes.* Here, we will describe properties of sums of independent random variables that have quickly decaying tails.

DEFINITION 2.1. Let f be a functions defined on a probability space (Ω, μ). For $1 \le \alpha \le 2$, define the α-Orlicz norm by

$$\|f\|_{\psi_\alpha} = \inf\left\{C > 0 : \mathbb{E} \exp\left(\frac{|f|^\alpha}{C^\alpha}\right) \le 2\right\}.$$

For basic facts regarding Orlicz norms, we refer the reader to [2, 19].

It is well known that a random variable has a bounded ψ_α norm for $1 \le \alpha \le 2$ if and only if it has a well behaved tail; that is, there is an absolute constant c such that for every $f \in L_{\psi_\alpha}$ and every $t \ge 1$,

$$\Pr(|f| \ge t) \le 2\exp(-ct^\alpha / \|f\|_{\psi_\alpha}^\alpha).$$

Conversely, there is an absolute constant c_1 such that if f displays a tail behavior dominated by $\exp(-t^\alpha/K^\alpha)$ for $1 \le \alpha \le 2$ then $\|f\|_{\psi_\alpha} \le c_1 K$.

There are several basic properties of sums of independent random variables we require. The proofs of these facts can be found, for example, in [2, 8, 19].

Note that if f has a subexponential tail then its empirical means concentrate around its true mean, with a tail behavior that is a mixture of subgaussian and subexponential. Indeed, the following result is a version of Bernstein's inequality and shows just that.

THEOREM 2.2. *There exists an absolute constant c for which the following holds. Let (Ω, μ) be a probability space and set $f : \Omega \to \mathbb{R}$ to be a function with a bounded ψ_1 norm. If $X_1, \dots, X_k$ are independent and distributed according to μ then for every $t > 0$,*

$$\Pr\left(\left| \frac{1}{k} \sum_{i=1}^k f(X_i) - \mathbb{E} f \right| \ge t \|f\|_{\psi_1} \right) \le 2\exp(-ck \min\{t^2, t\}).$$

If a function has mean zero and a bounded ψ_2 norm, one can obtain a purely subgaussian tail.

LEMMA 2.3. *There exists an absolute constant c for which the following holds. Let $Y_1, \dots, Y_k$ be independent random variables of mean zero. Then, for every $a_1, \dots, a_k \in \mathbb{R}$,*

$$\left\| \sum_{i=1}^k a_i Y_i \right\|_{\psi_2} \le c \left(\sum_{i=1}^k a_i^2 \|Y_i\|_{\psi_2}^2 \right)^{1/2}.$$

In particular, if $(X_i)_{i=1}^k$ are independent random variables distributed according to μ and f has zero mean, then for every $t \ge 1$,

$$\Pr\left(\left| \sum_{i=1}^k f(X_i) \right| \ge t k^{1/2} \|f\|_{\psi_2} \right) \le 2\exp(-c_1 t^2),$$

where c_1 is an absolute constant.

In what follows, we will assume that the ψ_2 and L_2 norms are equivalent on F in the following sense.

DEFINITION 2.4. A set $F \subset L_2(\mu)$ is L-subgaussian if $\|f\|_{\psi_2} \leq L\|f\|_{L_2}$ and $\|f-g\|_{\psi_2} \leq L\|f-g\|_{L_2}$ for every $f, g \in F$.

Next, let us turn to the definition of the $\gamma_{2,s}$ functionals [18]. Let (T,d) be a metric space. An *admissible sequence* of T is a sequence of subsets of T, $\{T_s\}_{s=0}^{\infty}$, such that $|T_0| = 1$ and for every $s \geq 1$, $|T_s| \leq 2^{2^s}$.

DEFINITION 2.5. For a metric space (T,d) and an integer $s_0 \geq 0$, let

$$\gamma_{2,s_0}(T,d) = \inf \sup_{t \in T} \sum_{s=s_0}^{\infty} 2^{s/2} d(t, T_s),$$

where the infimum is taken with respect to all admissible sequences of T. Set $\gamma_2(T,d) = \gamma_{2,0}(T,d)$.

Let $\pi_s : T \to T_s$ be a metric projection function onto T_s, that is, $\pi_s(t)$ is a nearest point to t in T_s with respect to the metric d. It is easy to verify that for every admissible sequence, every $t \in T$, and any $s_0 \geq 0$,

$$\sum_{s=s_0}^{\infty} 2^{s/2} d(\pi_{s+1}(t), \pi_s(t)) \leq (1 + 1/\sqrt{2}) \sum_{s=s_0}^{\infty} 2^{s/2} d(t, T_s)$$

and that the diameter of T satisfies $\operatorname{diam}(T,d) \leq 2\gamma_2(T,d)$. Moreover, it is clear that the $\gamma_{2,s}$ functionals are decreasing in s and are subadditive in T in the following sense. Let X be a normed space and consider two sets $A, B \subset X$. If $A + B = \{a + b : a \in A, b \in B\}$ is the Minkowski sum of A and B, then for every integer s,

$$\gamma_{2,s+1}(A+B, d) \leq \gamma_{2,s}(A,d) + \gamma_{2,s}(B,d).$$

There is a close connection between the $\gamma_{2,s}$ functionals with respect to L_2 norms and properties of Gaussian processes (see [3, 18] for expositions on these connections). Indeed, let $\{G_t : t \in T\}$ be a centered Gaussian process indexed by a set T and for every $s, t \in T$ define a metric on T by $d^2(s,t) = \mathbb{E}|G_s - G_t|^2$. One can show that under mild measurability assumptions on T,

$$c_1 \gamma_2(T,d) \leq \mathbb{E} \sup_{t \in T} G_t \leq c_2 \gamma_2(T,d),$$

where c_1 and c_2 are absolute constants. The upper bound is due to Fernique [4] and the lower bound is Talagrand's Majorizing Measures theorem [17]. The proof of both parts can be found in [18]. Thus, the γ_2 functional is finite if and only if the Gaussian process indexed by T is bounded.

Note that if $T \subset \mathbb{R}^n$ and $G_t = \sum_{i=1}^n g_i t_i$ then $d(u,t) = |u-t|$ and therefore

$$c_1 \gamma_2(T, |\cdot|) \le \mathbb{E} \sup_{t \in T} \sum_{i=1}^n g_i t_i \le c_2 \gamma_2(T, |\cdot|). \tag{2.1}$$

Just like $\gamma_2(T, L_2(\mu))$ determines the supremum of the canonical Gaussian process indexed by $T \subset L_2(\mu)$ (which we will always assume to satisfy the necessary measurability assumptions), the continuity of that process is determined by properties of the sequence $\gamma_{2,s}$.

Definition 2.6. Let $F \subset L_2(\mu)$ be a class of mean zero functions. Set $\{G_f : f \in F\}$ to be the centered Gaussian process indexed by F with a covariance structure endowed by $L_2(\mu)$; that is, for every $f, g \in F$, $\mathbb{E} G_f G_g = \langle f, g \rangle_{L_2(\mu)}$. We say that F is μ-pregaussian if it has a version with all sample functions bounded and uniformly continuous with respect to the $L_2(\mu)$ metric.

Theorem 2.7 [17, 18]. *Let $\{G_t : t \in T\}$ be a centered Gaussian process and endow T with the L_2 metric given by the covariance structure of the process, as above. Under measurability assumptions, the following are equivalent:*

1. *The map $t \to G_t(\omega)$ is uniformly continuous on T with probability* 1.
2. $\lim_{\delta \to 0} \mathbb{E} \sup_{d(u,t) \le \delta} |G_u - G_t| = 0$.
3. *There exists an admissible sequence of T such that*

$$\lim_{s_0 \to \infty} \sup_{t \in T} \sum_{s=s_0}^{\infty} 2^{s/2} d(t, \pi_s(t)) = 0.$$

In other words, T is pregaussian if and only if $\lim_{s \to \infty} \gamma_{2,s}(T, L_2) = 0$.

Remark 2.8. Theorem 2.7 is not proved in [18] but only stated there, and its formulation in [17] was done using the notion of majorizing measures rather than with the $\gamma_{2,s}$ functionals. Since the proof of the continuity theorem follows from an application of the Majorizing Measures theorem and since the latter is proved in [18] using the language of the γ_2-functional, it is not difficult to convert the proof of the continuity theorem from [17] and obtain Theorem 2.7. Moreover, as shown in [17], there is a quantitative connection between the modulus of continuity of $\{G_t : t \in T\}$ and the sequence $(\gamma_{2,s}(T, L_2))_{s=0}^{\infty}$. Since we will not use this quantitative estimate here, we will not formulate it.

Finally, let us define the covering and packing numbers of a metric space.

Definition 2.9. Let (T,d) be a metric space. The covering number of T at scale $\varepsilon>0$ with respect to the metric d is the smallest number of open balls of radius ε needed to cover T, and is denoted by $N(\varepsilon,T,d)$.

We set $e_k(T,d)=\inf\{\varepsilon : N(\varepsilon,T,d)\le 2^k\}$. $(e_k)_{k=0}^\infty$ are called the entropy numbers of T.

A set $A\subset T$ is called ε-separated if the distance between any two of its elements is at least ε. We denote by $D(\varepsilon,T,d)$ the cardinality of a maximal ε-separated subset of T.

It is standard to verify that for every $\varepsilon>0$, $N(\varepsilon,T,d)\le D(\varepsilon,T,d)\le N(\varepsilon/2,\,T,d)$, and thus one can use either one of the two notions freely.

3. The discrepancy of subsets of $\mathbb{R}^n$. We begin this section with a technical lemma which is at the heart of the proof of Theorem A. The lemma allows one to find a good choice of signs on roughly half of the coordinates, while the error incurred by the choice of coordinates and signs can be controlled using the geometric structure of T.

A preliminary result we need has to do with Bernoulli processes—the well-known Höffding inequality (see, e.g., [8, 19]).

Theorem 3.1. *Let $(\varepsilon_i)_{i=1}^n$ be independent, symmetric, $\{-1,1\}$-valued random variables. Then, for every $a\in\mathbb{R}^n$ and every $t>0$,*

$$\Pr\left(\sum_{i=1}^n \varepsilon_i a_i \ge t|a|\right)\le \exp(-t^2/2).$$

In particular,

$$\Pr\left(\left|\sum_{i=1}^n \varepsilon_i a_i\right| \ge t|a|\right)\le 2\exp(-t^2/2).$$

Let us formulate the main lemma.

Lemma 3.2. *Let*

$$\Phi(t)=\begin{cases}\log(e/t), & \text{if } 0<t\le 1,\\ t\exp(-t+1), & \text{if } t>1.\end{cases}$$

There exist absolute constants κ_1 and κ_2 for which the following holds. Assume that $(\lambda_s)_{s=1}^\infty$ is an increasing positive sequence tending to infinity, $(Q_s)_{s=1}^\infty$ is a positive sequence and n is an integer such that

$$\kappa_1\sum_{s=1}^\infty \lambda_s\Phi((\kappa_2 Q_s)^2)\le \frac{n}{100}.$$

Let $T\subset\mathbb{R}^n$ for which $0\in T$, set $(T_s)_{s=1}^\infty$ to be a sequence of subsets of T and let $T_0=\{0\}$. Consider maps $\pi_s : T\to T_s$ that satisfy that:

(a) *for every* $s \geq 1$, $|\{\pi_s(t) - \pi_{s-1}(t) : t \in T\}| \leq \lambda_s$,
(b) *for every* $t \in T$, $\lim_{s\to\infty} \pi_s(t) = t$.

Then, there exists $(\eta_i)_{i=1}^n \in \{-1,0,1\}^n$ *such that* $n/4 \leq |\{i : \eta_i = 0\}| \leq 3n/4$, *and for every* $t \in T$,

$$\left|\sum_{i=1}^n \eta_i t_i\right| \leq \sum_{s=1}^\infty Q_s |\pi_s(t) - \pi_{s-1}(t)|.$$

The proof is a combination of a chaining argument and the entropy method, which is frequently used in Discrepancy Theory (see, e.g., [1, 10, 16]). In the chaining mechanism, one takes the sets T_s to be finer and finer approximations of the set T and $\pi_s(t)$ is a nearest element to t in T_s with respect to the underlying metric (which is, in our case, the ℓ_2^n metric).

Recall that the entropy of a discrete random variable X taking values in a countable set Ω is

$$H(X) = -\sum_{\omega\in\Omega} p_\omega \log_2 p_\omega,$$

where $p_\omega = \Pr(X = \omega)$. The entropy function $H(X)$ indicates how close X is to being equally distributed; the more equally distributed X is, the larger $H(X)$ is.

The three facts we will need regarding the entropy are well known and we omit their proofs. First, if $H(X) \leq K$ then there is a value of X that is attained with probability at least 2^{-K}. Second, if X attains at most k values then $H(X) \leq \log_2 k$, and finally, if $X = (X_1, \ldots, X_m)$ is a random vector then $H(X) \leq \sum_{i=1}^m H(X_i)$.

In the entropic argument we will use, each "link" in each chain in T is assigned a random variable $X_\alpha : \{-1,1\}^n \to \mathbb{R}$ that depends on the link and on the chain. The idea is to show that with probability at least $2^{-\eta n}$, for every α, each random variable X_α falls in an interval I_α whose length is at most $Q_\alpha \|X_\alpha\|_{L_2}$. One would like to make these scaling factors Q_α as small as possible while still ensuring that conditions 1 and 2 hold, since those conditions imply that the intersection of the level sets of all the random variables X_α has the desired measure.

More details on the way entropic arguments have been used in the context of Discrepancy Theory may be found in [1, 11].

Before presenting the proof, one should mention that a chaining argument was implicit in Matoušek's result on the discrepancy of a subset of $\{0,1\}^n$ with a bounded VC dimension [10, 11].

The first step in the proof of Lemma 3.2 is the following entropy estimate. We denote by $[x]$ the integer value of x.

LEMMA 3.3. *There exists an absolute constant c for which the following holds. Let $a \in \mathbb{R}^n$, set $Z_a = \sum_{i=1}^n \varepsilon_i a_i$ and put*

$$W_a = \mathrm{sgn}(Z_a)[|Z_a|].$$

Then

$$-\sum_{i=-\infty}^{\infty} \Pr(W_a = i) \log \Pr(W_a = i) \le c\Phi(1/2|a|^2).$$

PROOF. By Höffding's inequality (Theorem 3.1), for every $j \in \mathbb{Z} \setminus \{0\}$,

$$p_j = \Pr(W_a = j) \le \Pr(Z_a \ge |j|) < \exp(-j^2/2|a|^2).$$

Also,

$$p_0 = \Pr(W_a = 0) = \Pr(-1 < Z_a < 1),$$

implying that

$$1 - p_0 = \Pr\left(\left|\sum_{i=1}^n \varepsilon_i a_i\right| \ge 1\right) \le 2\exp(-1/2|a|^2).$$

Consider $j \in \mathbb{Z}$ for which $|j| \ge \sqrt{2}|a|$. Since $f(x) = -x \log x$ is increasing in $[0, 1/e]$, it follows that for such values of j,

$$-p_j \log p_j \le \frac{j^2}{2|a|^2} \exp(-j^2/2|a|^2).$$

Fix an integer k which satisfies that $k \ge \sqrt{2}|a|$ and which will be named later, and observe that if we set $S = \sum_{1 \le |j| \le k} p_j$ then

$$-\sum_{1 \le |j| \le k} p_j \log p_j \le -\sum_{1 \le |j| \le k} \frac{S}{2k} \log(S/2k) = S \log(2k/S).$$

Clearly, $S \le 1 - p_0 \le 2\exp(-1/2|a|^2)$, and thus, if $\exp(-1/2|a|^2) \le 1/e$ (i.e., if $\sqrt{2}|a| \le 1$), then

$$S \log(2k/S) = \log k + 2(S/2)\log(2/S) \le \log k + \frac{1}{2|a|^2}\exp(-1/2|a|^2).$$

Otherwise, $S \log(2k/S) \le \log k + \frac{1}{e}\log(2e) \le 1 + \log k$, implying that

$$-\sum_{1 \le |j| \le k} p_j \log p_j \le \log k + \begin{cases} \dfrac{1}{2|a|^2}\exp(-1/2|a|^2), & \text{if } \sqrt{2}|a| \le 1, \\ 1, & \text{otherwise.} \end{cases}$$

Moreover,

$$\begin{aligned} -\sum_{|j|\geq k+1} p_j \log p_j &\leq 2 \sum_{j\geq k+1} \frac{j^2}{2|a|^2} \exp(-j^2/2|a|^2) \\ &\leq 2\int_k^\infty \frac{x^2}{2|a|^2}\exp(-x^2/2|a|^2) \\ &\leq (k+2|a|)\exp(-k^2/2|a|^2). \end{aligned}$$

Therefore,

$$\begin{aligned} -\sum_{j=\infty}^{\infty} p_j \log p_j &= -\sum_{|j|>k} p_j \log p_j - p_0 \log p_0 - \sum_{1\leq|j|\leq k} p_j \log p_j \\ &\leq (k+2|a|)\exp(-k^2/2|a|^2) + 2\exp(-1/2|a|^2) \\ &\quad + \log k + \begin{cases} \dfrac{1}{2|a|^2}\exp(-1/2|a|^2), & \text{if } \sqrt{2}|a|\leq 1, \\ 1, & \text{otherwise.} \end{cases} \end{aligned}$$

Now, consider the following three cases. First, if $\sqrt{2}|a|\leq 1$, take $k=1$, and thus

$$-\sum_{j=\infty}^{\infty} p_j \log p_j \leq \frac{c_1}{|a|^2}\exp(-1/2|a|^2).$$

If $1<\sqrt{2}|a|\leq e$ set k to be a suitable absolute constant and if $\sqrt{2}|a|>e$, put $k\sim|a|\log(\sqrt{2}|a|)$. Therefore, in both these cases

$$-\sum_{j=\infty}^{\infty} p_j \log p_j \leq c_2 \log(\sqrt{2}|a|),$$

and our claim follows. $\square$

Proof of Lemma 3.2. Without loss of generality, assume that T is finite. Recall that $0\in T$ and that $T_0=\{0\}$, consider the sets T_s and the maps $\pi_s : T\to T_s$, let $\Delta_s(t)=\pi_s(t)-\pi_{s-1}(t)$ and put $\Delta_s=\{\pi_s(t)-\pi_{s-1}(t) : t\in T\}$. Let $(\lambda_s)_{s=1}^\infty$ and $(Q_s)_{s=1}^\infty$ be as in the assumptions of the lemma and set $(\varepsilon_i)_{i=1}^n$ to be independent, symmetric, $\{-1,1\}$-valued random variables.

Consider the Bernoulli process $t\to Z_t=\sum_{i=1}^n \varepsilon_i t_i$. Since Z_t is linear in t and $\pi_0(t)=0$, then for every $t\in T$,

$$t=\sum_{s=1}^\infty \Delta_s(t) \quad \text{and} \quad Z_t=\sum_{s=1}^\infty Z_{\pi_s(t)}-Z_{\pi_{s-1}(t)}=\sum_{s=1}^\infty Z_{\Delta_s(t)}.$$

For every $s \geq 1$ and $u \in \Delta_s$ define

$$\tilde{W}_{u,s} = \frac{Z_u}{|u|Q_s}, \qquad W_{u,s} = \operatorname{sgn}(\tilde{W}_{u,s})[|\tilde{W}_{u,s}|].$$

Observe that $(W_{u,s})_{u \in \Delta_s, s=1,2,\ldots}$ is a vector that takes a finite number of values. Since the entropy is subadditive then

$$H((W_{u,s}) : u \in \Delta_s, s = 1, 2, \ldots) \leq \sum_{s=1}^{\infty} \sum_{u \in \Delta_s} H(W_{u,s}) = (*).$$

Suppose that one can find $(Q_s)_{s=1}^{\infty}$ for which $(*) \leq n/100$. By the properties of the entropy, this implies that there are numbers $\{\ell_{u,s} \in \mathbb{Z} : u \in \Delta_s, s = 1, 2, \ldots\}$ such that

$$\Pr((\varepsilon_i)_{i=1}^n : \forall u \in \Delta_s, s \geq 1, W_{u,s} = \ell_{u,s}) \equiv \Pr(A) \geq 2^{-n/100}. \tag{3.1}$$

Since $|A| \geq 2^{0.99n}$, there will be at least two vectors $(\varepsilon_i)_{i=1}^n$ and $(\varepsilon_i')_{i=1}^n$ in A that differ on at most $3n/4$ coordinates and on at least $n/4$ of them. The desired sequence will then be $(\eta_i)_{i=1}^n = (\frac{\varepsilon_i - \varepsilon_i'}{2})_{i=1}^n$. Indeed, for $u \in \Delta_s$,

$$\left| \sum_{i=1}^n \eta_i u_i \right| = \frac{1}{2} \left| \sum_{i=1}^n \varepsilon_i u_i - \sum_{i=1}^n \varepsilon_i' u_i \right| \leq Q_s |u|,$$

implying that every $t \in T$ satisfies

$$\left| \sum_{i=1}^n \eta_i t_i \right| = \left| \sum_{s \geq 1} \sum_{i=1}^n \eta_i (\Delta_s(t))_i \right| \leq \sum_{s=1}^{\infty} Q_s |\Delta_s(t)|.$$

Hence, to complete the proof, it remains to show that for a sequence $(Q_s)_{s=1}^{\infty}$ that satisfies the assumptions of the lemma, $(*) \leq n/100$. Applying Lemma 3.3 for $a = u/|u|Q_s$, and since $1/2|a|^2 = Q_s^2/2$, it is evident that $H(W_{u,s}) \lesssim \Phi(Q_s^2/2)$. Thus,

$$\sum_{s=1}^{\infty} \sum_{u \in \Delta_s} H(W_{u,s}) \lesssim \sum_{s=1}^{\infty} |\Delta_s| \sup_{u \in \Delta_s} H(W_{u,s}) \lesssim \sum_{s=1}^{\infty} \lambda_s \Phi(Q_s^2/2),$$

proving our claim. $\square$

We will apply Lemma 3.2 in two typical situations. The first case will lead to a bound on the discrepancy of a set using the $\gamma_{2,s}$ functionals of the set and of its coordinate projections. The second will result is an entropy integral type bound, presented in Section 3.1, which will then be used to re-prove Spencer's result on the discrepancy of a finite set system [11, 16] and Matoušek's VC theorem [10, 11].

Corollary 3.4 below will play a central part in the proof of Theorem A. Since it follows from a simple computation, we omit its proof.

COROLLARY 3.4. *There exist absolute constants κ_3, κ_4, κ_5 for which the following holds. Let $T \subset \ell_2^n$, assume that $0 \in T$, set $s_n = \max\{s : 2^{2^{s+1}} \leq \kappa_3 n\}$ and put T_s to be a collection of subsets of T with $|T_s| \leq 2^{2^s}$. Then, if*

$$Q_s = \kappa_4 \begin{cases} \exp(-\kappa_5 n^{1/2}), & \text{if } s < s_n, \\ 1, & \text{if } s = s_n, \\ 2^{s/2}, & \text{if } s > s_n, \end{cases}$$

there exists $(\eta_i)_{i=1}^n \in \{-1,0,1\}^n$ such that $n/4 \leq |\{i : \eta_i = 0\}| \leq 3n/4$, and for every $t \in T$,

$$\left|\sum_{i=1}^n \eta_i t_i\right| \leq \sum_{s=1}^\infty Q_s |\pi_s(t) - \pi_{s-1}(t)|,$$

where $\pi_s(t)$ is a nearest point to t in T_s.

3.1. *An entropy integral argument.* In this section, we will prove an analog of Dudley's entropy integral bound (see, e.g., [8, 18]) in the context of discrepancy. The entropy integral is often used to upper bound $\sup_{t\in T}|\sum_{i=1}^n \varepsilon_i t_i|$ for a typical $(\varepsilon_i)_{i=1}^n$, but here we will present a modified version that allows one to control $\inf_\eta \sup_{t\in T}|\sum_{i=1}^n \eta_i t_i|$, where the infimum is taken with respect to all $\eta = (\eta_i)_{i=1}^n \in \{-1,0,1\}^n$ for which roughly half the coordinates are nonzero.

Let $T \subset \ell_2^n$ and recall that for every $\varepsilon > 0$, $D(\varepsilon) = D(\varepsilon, T, \ell_2^n)$ is the cardinality of a maximal ε-separated subset of T. Also, set

$$u(\varepsilon) = \begin{cases} \sqrt{\log\left(\dfrac{eD(\varepsilon)}{n}\right)}, & \text{if } D(\varepsilon) \geq n, \\ \exp\left(-\sqrt{\dfrac{n}{D(\varepsilon)}} + 1\right), & \text{if } D(\varepsilon) < n. \end{cases}$$

THEOREM 3.5. *There exist an absolute constant c for which the following holds. If $T \subset \ell_2^n$ and $0 \in T$, then there exist $(\eta_i)_{i=1}^n \in \{-1,0,1\}^n$, such that $n/4 \leq |\{i : \eta_i = 0\}| \leq 3n/4$ and for every $t \in T$,*

$$\left|\sum_{i=1}^n \eta_i t_i\right| \leq c \int_0^{\operatorname{diam}(T)} u(\varepsilon)\, d\varepsilon. \tag{3.2}$$

REMARK 3.6. Recall that Dudley's entropy integral bound shows that

$$\mathbb{E} \sup_{t \in T}\left|\sum_{i=1}^n \varepsilon_i t_i\right| \leq c_1 \int_0^{\operatorname{diam}(T)} \sqrt{\log D(\varepsilon)}\, d\varepsilon,$$

for a suitable absolute constant c_1. Clearly, this entropy integral may be considerably larger than the quantity we have in Theorem 3.5. It is also evident that if one could iterate Theorem 3.5 for the set $P_I T \subset \ell_2^{|I|}$, where $I = \{i : \eta_i = 0\}$, and continue in the same manner, then one would likely improve upon the bound resulting from the standard entropy integral bound that holds for a typical choice of signs, if indeed distances in $P_I T$ shrink relative to distances in T.

The proof of Theorem 3.5 is based on Lemma 3.2. It requires two additional simple results. Since their proofs are standard, we shall not present them here.

LEMMA 3.7. *There exist absolute constants c_1, c_2, c_3 and c_4 for which the following holds. Let $T \subset \ell_2^n$, set ν_n to be the largest integer s satisfying $2^s \le c_1 n$ and define*

$$\lambda_s = \begin{cases} c_2 2^s, & \text{if } s \le \nu_n, \\ c_3 n 2^{2^{s-\nu_n}-1}, & \text{if } s > \nu_n. \end{cases}$$

Then conditions (a) *and* (b) *of Lemma 3.2 hold if one selects*

$$Q_s = c_4 \begin{cases} \exp(-2 \cdot 2^{(s-\nu_n)/2}), & \text{if } s \le \nu_n, \\ 2^{(s-\nu_n)/2}, & \text{if } s > \nu_n. \end{cases}$$

LEMMA 3.8. *Let g and f be nonincreasing, nonnegative functions and let $(\varepsilon_s)_{s=0}^m$ be a decreasing sequence. If for every $s \ge 1$, $g(\varepsilon_{s-1}) \ge f(\varepsilon_s)$, and if there is $\alpha > 0$ such that for every $s \ge 1$, $f(\varepsilon_s) - f(\varepsilon_{s-1}) \ge \alpha f(\varepsilon_s)$ then*

$$\int_{\varepsilon_m}^{\varepsilon_0} g(\varepsilon)\, d\varepsilon + \varepsilon_m f(\varepsilon_m) \ge \alpha \sum_{s=1}^m f(\varepsilon_s)\varepsilon_{s-1}.$$

PROOF OF THEOREM 3.5. Let $(\lambda_s)_{s=1}^\infty$ and $(Q_s)_{s=1}^\infty$ be as in Lemma 3.7. Without loss of generality assume that T is a finite set and define the sets T_s iteratively, as follows. Set m to be the first integer such that $|T| \le \lambda_m$, let $T_s = T$ for $s \ge m$ and set $\varepsilon_m = 0$. For $m-1$, let $\varepsilon_{m-1} = \inf\{\varepsilon : D(\varepsilon, T_m, \ell_2^n) \le \lambda_{m-1}\}$ and put T_{m-1} to be a maximal ε_{m-1}-separated subset of T_m whose cardinality is at most λ_{m-1}. Continue in this way to construct the sets T_s for $s = m-1, \dots, 1$. For every s, let $\pi_s(t)$ be a nearest point to $\pi_{s+1}(t)$ in T_s.

Let $s \le m$, and since the sets T_s are nested, then $|\{\pi_s(t) - \pi_{s-1}(t) : t \in T\}| \le |T_s|$ and $|\pi_s(t) - \pi_{s-1}(t)| \le \varepsilon_{s-1}$ for every $t \in T$. Therefore, applying Lemmas 3.2 and 3.7, there is a choice $(\eta_i)_{i=1}^n \in \{-1, 0, 1\}^n$ with $n/4 \le |\{i : \eta_i = 0\}| \le 3n/4$ such that for every $t \in T$,

$$\left|\sum_{i=1}^n \eta_i t_i\right| \le c_1 \left(\sum_{s \le \nu_n} \exp(-2 \cdot 2^{(\nu_n - s)/2})\varepsilon_{s-1} + \sum_{s > \nu_n}^m 2^{(s-\nu_n)/2}\varepsilon_{s-1}\right). \tag{3.3}$$

It remains to bound the sums in (3.3) by the appropriate integrals, using Lemma 3.8. First, for $s > \nu_n$ let

$$f(\varepsilon) = \sum_{s=\nu_n+1}^{m} 2^{(s-\nu_n)/2} \mathbb{1}_{(\varepsilon_{s+1},\varepsilon_s]}, \qquad g(\varepsilon) = \sum_{s=\nu_n+1}^{m} 2^{(s-\nu_n)/2} \mathbb{1}_{(\varepsilon_s,\varepsilon_{s-1}]}.$$

Clearly, in $[\varepsilon_m, \varepsilon_{\nu_n}] = [0, \varepsilon_{\nu_n}]$, f and g are nonincreasing and nonnegative, for every ε in that range

$$f(\varepsilon) \le g(\varepsilon) \le \sqrt{2} f(\varepsilon) \lesssim u(\varepsilon),$$

and the conditions of Lemma 3.8 hold. Since $\varepsilon_m = 0$, then

$$\sum_{s>\nu_n}^{m} 2^{(s-\nu_n)/2} \varepsilon_{s-1} \lesssim \int_0^{\varepsilon_n} u(\varepsilon).$$

For the other term in (3.3), if $s \le \nu_n$ then $2^s \sim \lambda_s \le D(\varepsilon_s)$ and the sum is estimated in a similar way. $\square$

3.1.1. *Spencer's theorem.* Let us show how Theorem 3.5 can be used to prove a version of Spencer's celebrated result from [16] (see also [1, 11]).

THEOREM 3.9. *There exists an absolute constant c such that if $T \subset B_\infty^n$ is of cardinality $m \ge n$, then*

$$\mathrm{disc}(T) \le c\sqrt{n \log\left(\frac{em}{n}\right)}.$$

PROOF. Without loss of generality, assume that $0 \in T$. Using the notation of Theorem 3.5, for every $\varepsilon > 0$, $u(\varepsilon) \le \sqrt{\log(em/n)}$, and since $T \subset B_\infty^n$ then $\mathrm{diam}(T) \le \sqrt{n}$. Hence, there are $(\eta_i)_{i=1}^n \in \{-1, 0, 1\}^n$ for which $n/4 \le |\{i : \eta_i = 0\}| \le 3n/4$ and for every $t \in T$,

$$\left|\sum_{i=1}^n \eta_i t_i\right| \le c_1 \int_0^{\sqrt{n}} \sqrt{\log(em/n)}\, d\varepsilon \le c_1 \sqrt{n \log(em/n)}.$$

Now the result follows by repeating this argument for $P_{I_1} T$, where $I_1 = \{i : \eta_i = 0\}$, an so on. $\square$

3.1.2. *Matoušek's VC theorem.* A well-known measure of complexity for subsets of $\{0,1\}^n$ is the VC dimension of the set (its real value counterpart will be used in Section 6).

DEFINITION 3.10. Let $T \subset \{0,1\}^n$. We say that $\sigma \subset \{1,\dots,n\}$ is shattered by T if $P_\sigma T = \{0,1\}^\sigma$—that is, if the coordinate projection $P_\sigma T = \{(t_i)_{i\in\sigma} : t \in T\}$ is the entire combinatorial cube on these coordinates. Define $\mathrm{VC}(T)$ to be the maximal cardinality of a subset of $\{1,\dots,n\}$ that is shattered by T.

In [10], Matoušek proved that the discrepancy of a VC class is polynomially better than could be expected from a random choice of signs. He obtained the best possible estimate for the discrepancy of VC-subsets of $\{0,1\}^n$ as a function of the dimension n.

THEOREM 3.11. *For every integer d, there is a constant $c(d)$ for which the following holds. If $T \subset \{0,1\}^n$ and $\mathrm{VC}(T) \le d$, then $\mathrm{disc}(T) \le c(d)n^{1/2-1/2d}$.*

To prove Matoušek's theorem, recall the following fundamental property of a VC class, due to Haussler [7].

LEMMA 3.12. *If $T \subset \{0,1\}^n$ and $\mathrm{VC}(T) \le d$, then for every $I \subset \{1,\dots,n\}$ and every $0 < \varepsilon \le |I|^{1/2}$,*

$$D(\varepsilon, P_I T, \ell_2^I) \le c(d)\left(\frac{|I|^{1/2}}{\varepsilon}\right)^{2d},$$

where $c(d)$ is a constant that depends only on d.

PROOF OF THEOREM 3.11. Again, we may assume that $0 \in T$ and view T as a subset of $\mathbb{R}^n$. Let $\varepsilon_n = \inf\{\varepsilon : D(\varepsilon) \le n\}$. Therefore, $\varepsilon_n \le c_1(d)n^{1/2-1/2d}$. A change of variables shows that

$$\int_0^{\varepsilon_n} \sqrt{\log(eD(\varepsilon)/n)}\, d\varepsilon \le c_2(d)n^{1/2-1/2d},$$

$$\int_{\varepsilon_n}^{\mathrm{diam}(T)} \exp(-\sqrt{n/D(\varepsilon)})\, d\varepsilon \le c_2(d)n^{1/2-1/2d}.$$

Hence, there is a choice of $(\eta_i^0)_{i=1}^n \in \{-1,0,1\}^n$ such that for every $t \in T$

$$\left|\sum_{i=1}^n \eta_i^0 t_i\right| \le c_3(d)n^{1/2-1/2d},$$

and if we set $I_1 = \{i : \eta_i^1 = 0\}$ then $|I_1| \le 3n/4$. Since $\mathrm{VC}(P_{I_1}T) \le d$ then repeating the same argument for the set $P_{I_1}T$, there are $(\eta_i^1)_{i\in I_1} \in \{-1,0,1\}^{I_1}$ such that for every $t \in T$, $|\sum_{i\in I_1} \eta_i^1 t_i| \le c_3(d)|I_1|^{1/2-1/2d}$, and so on. Therefore, there is a choice of signs $(\varepsilon_i)_{i=1}^n$ such that for every $t \in T$,

$$\left|\sum_{i=1}^n \varepsilon_i t_i\right| \le c_3(d)\sum_j |I_j|^{1/2-1/2d} \le c_4(d)n^{1/2-1/2d},$$

where we have used the fact that for every j, $|I_j| \le 3|I_{j-1}|/4$. $\square$

The proof of Theorem 3.11 illustrates once again the main property we used to bound the discrepancy of a subset of $\mathbb{R}^n$. It is not enough for the set to be small in the sense of its metric entropy; what is needed is additional control on the "size" of all of the set's coordinate projections. One way of controlling those coordinate projections is by taking into account information about the position of vectors in the set, since coordinate projections of vectors in a good position shrink norms and mutual distances.

4. A decomposition theorem for subgaussian processes. It is clear from our estimate on the discrepancy of a set $T \subset \mathbb{R}^n$ that it would be useful to control the ℓ_2^I distances between points in T for every $I \subset \{1,\dots,n\}$—that is, distances between coordinate projections of elements of T. One would be able to obtain a good bound on $\mathrm{disc}(T)$ if T is not too rich and if for every $I \subset \{1,\dots,n\}$ and every $x, y \in T$, $\|x-y\|_{\ell_2^I}$ is significantly smaller than $\|x-y\|_{\ell_2^n}$. Unfortunately, usually this is not true even for a single vector $z = x-y$. Indeed, if z is supported in I then $\|z\|_{\ell_2^I}$ does not "shrink" at all. On the other hand, if the coordinates of z are roughly equal, then the coordinate projection onto any I shrinks the norm of z by a factor of $(|I|/n)^{1/2}$.

It is well known that a strong shrinking phenomenon is exhibited by vectors in a general position. In this section, we will show that if a class of functions F is L-subgaussian, then a shrinking phenomenon happens for a typical set

$$T = P_\sigma F = \{(f(X_i))_{i=1}^k : f \in F\},$$

uniformly for all coordinate projections of T.

4.1. *Shrinking for a single function.* As a starting point, let us describe the so-called "standard shrinking" phenomenon for a single function f. Let f be a function for which $\|f\|_{\psi_2} \le L\|f\|_{L_2}$. Then, concentration implies that with high probability,

$$\|f\|_{L_2^k} = \left(\frac{1}{k}\sum_{i=1}^k f^2(X_i)\right)^{1/2} \sim \|f\|_{L_2}.$$

However, as we mentioned above, the shrinking phenomenon one needs here is more general—that for every subset $I \subset \{1,\dots,k\}$, the L_2^I norm of f is upper bounded (possibly up to a logarithmic factor) by $\|f\|_{L_2}$ (which translates in the ℓ_2 normalization to the shrinking of the norm). The following lemma shows that this stronger claim is true as well whenever f is L-subgaussian.

LEMMA 4.1. *For every $0<\delta<1$ and $L>0$, there is a constant $c(\delta,L)$ for which the following holds. If $\|f\|_{\psi_2}\le L\|f\|_{L_2}$ then for every integer k, with probability at least $1-\delta$, for every $I\subset\{1,\dots,k\}$,*

$$\|f\|_{L_2^I}\le c(\delta,L)\sqrt{\log(ek/|I|)}\|f\|_{L_2}.$$

PROOF. Fix k and $I\subset\{1,\dots,k\}$. Since $\|f^2\|_{\psi_1}=\|f\|_{\psi_2}^2$, then by Bernstein's inequality, for every $t>0$,

$$\Pr\left(\left|\frac{1}{|I|}\sum_{i\in I}f^2(X_i)-\mathbb{E}f^2\right|\ge t\|f\|_{\psi_2}^2\right)\le 2\exp(-c_0|I|\min\{t^2,t\}).$$

Let $m\le c_1k$ and recall that there are at most $(ek/m)^m$ subsets of $\{1,\dots,k\}$ of cardinality m. Hence, it suffices to take $t=\beta(\delta)\log(ek/m)\ge 1$ and obtain that with probability of at least $1-2\exp(-c_0\beta m\log(ek/m))$, for every subset I of $\{1,\dots,k\}$ of cardinality m,

$$\|f\|_{L_2^I}=\left(\frac{1}{m}\sum_{i\in I}f^2(X_i)\right)^{1/2}\le c_2(\delta)L\sqrt{\log(ek/m)}\|f\|_{L_2}. \tag{4.1}$$

Therefore, summing the probabilities with respect to m, it follows that for the correct choice of β, with probability at least $1-\delta$, (4.1) is true for all subsets of $\{1,\dots,k\}$ of cardinality at most c_1k. The claim now easily follows. $\square$

4.2. *Shrinking for a class of functions.* When one attempts to generalize this simple shrinking argument to a class of functions, one faces a problem: the probabilistic estimate obtained in the proof of Lemma 4.1 does not allow one to control many functions simultaneously. Thus, a naive extension of that result is simply too weak to lead to a function class analog of the shrinking phenomenon.

To formulate the shrinking phenomenon for an L-subgaussian class of functions, let us recall some notation. For any two sets A and B in a vector space, $A+B=\{a+b:a\in A,b\in B\}$, and for a class of functions F, a random sample $\sigma=(X_1,\dots,X_k)$ and $I\subset\{1,\dots,k\}$,

$$P_\sigma F=\{(f(X_i))_{i=1}^k:f\in F\},\qquad P_I^\sigma F=\{(f(X_i))_{i\in I}:f\in F\}.$$

For every integer m, let $W_m=\{x\in\mathbb{R}^m:x_j^*\le 1/\sqrt{j},j=1,\dots,m\}$, where $(x_j^*)_{j=1}^m$ is a monotone nonincreasing rearrangement of $(|x_j|)_{j=1}^m$. Thus, W_m is the unit ball of the weak ℓ_2 space $\ell_{2,\infty}^m$. Denote by V_m the collection of all subsets of $\{1,\dots,k\}$ of cardinality at most m and set τ_m to be the smallest integer s such that $2^{2^s}\ge\exp(m\log(ek/m))\ge|V_m|$.

THEOREM 4.2. *For every $0 < \delta < 1$ and $L > 0$ there exist constants c_1, c_2, c_3 and k_0 depending only on L and δ for which the following holds. Let F be an L-subgaussian class of functions and assume that for each $f \in F$, $\mathbb{E}f = \alpha$ for some $\alpha \in \mathbb{R}$. Then, for every integer k and every $m \le k$, there are sets F_1^m and $F_2^m \subset F$ with the following properties. First, $F \subset F_1^m + F_2^m$; second, with μ^k-probability of at least $1 - \delta$, if $\sigma = (X_1, \dots, X_k)$ then:*

1. *For every integer $m \le k$ and every $I \subset \{1, \dots, k\}$ of cardinality m,*
$$P_I^\sigma F_1^m \subset c_1 \gamma_{2,\tau_m}(F, L_2) W_m.$$
2. *For every $f, h \in F_2^m$ and every $I \subset \{1, \dots, k\}$ of cardinality m,*
$$\|f - h\|_{L_2^I} \le c_2 \sqrt{\log(ek/m)} \|f - h\|_{L_2}.$$
3. *If $k \ge k_0$ then for every $m \le c_3 k$ and every $f, h \in F_2^m$,*
$$\|f - h\|_{L_2} \le \sqrt{2} \|f - h\|_{L_2^\sigma}.$$

The way Theorem 4.2 should be understood is as follows. Consider a typical $\sigma = (X_1, \dots, X_k)$ and let $T = P_\sigma F \subset \ell_2^k$. Then, for every $I \subset \{1, \dots, k\}$ the further coordinate projection satisfies $P_I T \subset P_I T_1 + P_I T_2$ where $T_1, T_2 \subset \ell_2^k$ depend only on the cardinality of I and not on I itself, and $T_2 \subset T$. The set $P_I T_1$ captures the "peaky" part of $P_I T$ and is contained in a relatively small set: a ball in $\ell_{2,\infty}$ whose radius depends on the "complexity" of the class F. The set T_2 consists of vectors that satisfy the desired shrinking property. Indeed, for every $(f(X_i))_{i=1}^k, (h(X_i))_{i=1}^k \in T_2$ and every $I \subset \{1, \dots, k\}$ of cardinality m one has

$$\begin{aligned}\left(\sum_{i \in I} (f(X_i) - h(X_i))^2\right)^{1/2} &\le c_1 \sqrt{m \log(ek/m)} \|f - h\|_{L_2} \\ &\le c_2 \sqrt{\frac{m}{k} \log(ek/m)} \left(\sum_{i=1}^k (f(X_i) - h(X_i))^2\right)^{1/2},\end{aligned}$$

where the last inequality holds if $m \le c_3 k$.

PROOF OF THEOREM 4.2. Fix an integer k. For every integer $m \le k$, let $(H_{s,m})_{s=\tau_m}^\infty$ be an almost optimal admissible sequence of F with respect to $\gamma_{2,\tau_m}(F, \psi_2)$, and set π_s^m to be the metric projection onto $H_{s,m}$ with respect to the ψ_2 norm. For every such m we will construct two sets of functions, F_1^m and F_2^m such that $F \subset F_1^m + F_2^m$ as follows: let $F_1^m = \{f - \pi_{\tau_m}^m(f) : f \in F\}$ and set $F_2^m = \{\pi_{\tau_m}^m(f) : f \in F\}$ [and from here on we will omit the superscript m and write $\pi_s(f)$ instead of $\pi_s^m(f)$]. Note that this choice of decomposition depends only on m and does not depend on k.

For every $I \in V_m$ set $Z_f^I = \sum_{i \in I}(f(X_i) - \mathbb{E}f)$ and observe that

$$Z_f^I - Z_{\pi_{\tau_m}(f)}^I = \sum_{s > \tau_m} Z_{\pi_s(f)}^I - Z_{\pi_{s-1}(f)}^I = \sum_{s > \tau_m} \sum_{i \in I} (\pi_s(f) - \pi_{s-1}(f))(X_i),$$

since the expectation of all the functions in F is the same. Thus, for every $f \in F$, $\pi_s(f) - \pi_{s-1}(f)$ has mean zero, and for every $s > \tau_m$ and every $t \geq 1$,

$$\begin{aligned} &\Pr\left(\left|\sum_{i \in I}(\pi_s(f) - \pi_{s-1}(f))(X_i)\right| \geq t\|\pi_s(f) - \pi_{s-1}(f)\|_{\psi_2}\sqrt{|I|}\right) \\ &\quad \leq 2\exp(-c_0 t^2). \end{aligned}$$

Let $t = u2^{s/2}$ for $u \geq c_1$, where c_1 is a constant to be named later. Because of our choice of s, $|V_m| \leq 2^{2^s}$ and $|H_{s,m}| \cdot |H_{s-1,m}| \leq 2^{2^{s+1}}$, and thus

$$\begin{aligned} &\Pr(\exists f \in F, I \in V_m : |Z_{\pi_s(f)}^I - Z_{\pi_{s-1}(f)}^I| \geq u2^{s/2}\|\pi_s(f) - \pi_{s-1}(f)\|_{\psi_2}\sqrt{|I|}) \\ &\quad \leq 2^{2^{s+1}}|V_m| \cdot 2\exp(-c_0 u^2 2^s) \leq \exp(-c_2 u^2 2^s). \end{aligned}$$

Hence, summing over $s > \tau_m$, it follows that with probability at least

$$1 - \sum_{s > \tau_m} \exp(-c_2 u^2 2^s) \geq 1 - \exp(-c_3 u^2 2^{\tau_m}),$$

for every $f \in F$ and every $I \in V_m$

$$\left|\sum_{i \in I}(f - \pi_{\tau_m}(f))(X_i)\right| \leq u\sqrt{|I|} \sum_{s > \tau_m} 2^{s/2}\|\pi_s(f) - \pi_{s-1}(f)\|_{\psi_2}.$$

Summing the probabilities for all possible integers $1 \leq m \leq k$ and noting that for every $1 \leq m \leq k$, $2^{\tau_m} \gtrsim m\log(ek/m)$, it is evident that for $u \geq c_1$ there is a set $\mathcal{A} \subset \Omega^k$ with probability at least $1 - \exp(-c_4 u^2)$ for which the following holds. For every $(X_i)_{i=1}^k \in \mathcal{A}$, every $1 \leq m \leq k$, every $h \in F_1^m$ and every $I \in V_m$

$$\left|\sum_{i \in I} h(X_i)\right| \leq 2u\sqrt{|I|}\gamma_{2,\tau_m}(F, \psi_2),$$

where we have used the fact that $(H_{s,m})_{s=\tau_m}^{\infty}$ is an almost optimal admissible sequence with respect to $\gamma_{2,\tau_m}(F, \psi_2)$.

Fix $(X_1, \ldots, X_k) \in \mathcal{A}$, $1 \leq m \leq k$, $I \subset \{1, \ldots, k\}$ of cardinality m and $h \in F_1^m$. Consider the sets $I_+ = \{i \in I : h(X_i) \geq 0\}$ and $I_- = \{i \in I : h(X_i) < 0\}$ and note that both are in V_m. Since $x^{1/2}$ is increasing, then on the set $\mathcal{A}$

$$\sum_{i \in I} |h(X_i)| \leq 4u\sqrt{|I|}\gamma_{2,\tau_m}(F, \psi_2). \tag{4.2}$$

In particular, if $(h_i^*)_{i=1}^k$ is a nonincreasing rearrangement of $(|h(X_i)|)_{i=1}^k$ then by (4.2) applied to the set I_j consisting of the $j \le m$ largest elements of $(|h(X_i)|)_{i=1}^k$,

$$h_j^* \le \frac{1}{j}\sum_{i=1}^{j} h_i^* \le \frac{1}{j}\cdot 4uj^{1/2}\gamma_{2,\tau_m}(F,\psi_2) \le 4Lu\gamma_{2,\tau_m}(F,L_2)/\sqrt{j};$$

thus, $P_I^\sigma F_1^m \subset 4Lu\gamma_{2,\tau_m}(F,L_2)W_m$.

Turning our attention to the sets F_2^m, we will show that with high probability, for every $1 \le m \le k$ and every $I \subset \{1,\ldots,k\}$ of cardinality m, the coordinate projection $P_I:(F_2^m, L_2) \to (F_2^m, L_2^I)$ has a well behaved Lipschitz constant. To that end, fix $1 \le m \le k$, set $G_m = \{|f_1 - f_2| : f_i \in F_2^m\}$ and recall that for every function g, $\|g^2\|_{\psi_1} = \|g\|_{\psi_2}^2$. Hence, by Bernstein's inequality, for every $g \in G_m$ and every $t > 1$,

$$\Pr\left(\left|\frac{1}{m}\sum_{i=1}^m g^2(X_i) - \mathbb{E}g^2\right| \ge t\|g\|_{\psi_2}^2\right) \le 2\exp(-c_0 m \min(t^2,t)).$$

Let E_m be the collection of subsets of $\{1,\ldots,k\}$ of cardinality m. Since $|G_m| \le |F_2^m|^2 \le 2^{2^{\tau m+1}}$ and $|E_m| \le |V_m| \le 2^{2^{\tau m}}$, then by taking $u \ge c_5$ and $t = u\log(ek/m) \ge 1$,

$$\Pr\left(\exists g \in G_m, I \in E_m : \left|\frac{1}{m}\sum_{i\in I} g^2(X_i) - \mathbb{E}g^2\right| \ge \|g\|_{\psi_2}^2 u\log(ek/m)\right)$$
$$\le 2^{2^{\tau m+2}}\exp(-c_0 um\log(ek/m)) \le \exp(-c_6 um\log(ek/m)).$$

Summing over all possible $1 \le m \le k$, there is a subset $\mathcal{B} \subset \Omega^k$ of probability at least $1 - \exp(-c_7 u)$ on which the following holds. For every $1 \le m \le k$, every $f_1, f_2 \in F_2^m$ and every $I \in E_m$,

$$\begin{aligned}\|f_1 - f_2\|_{L_2^I}^2 &\le \|f_1 - f_2\|_{L_2}^2 + u\log(ek/m)\|f_1 - f_2\|_{\psi_2}^2 \\ &\le 2L^2 u\log(ek/m)\cdot\|f_1 - f_2\|_{L_2}^2.\end{aligned} \tag{4.3}$$

Thus, fix a "legal" choice of u for which $\Pr(\mathcal{A}\cap\mathcal{B}) \ge 1 - \delta/2$. Since both (4.2) and (4.3) hold on that event, the proof of the first and second claims is evident.

For the third part, fix $t < 1/2$ to be named later. Again, by Bernstein's inequality and since F is L-subgaussian, then with probability at least $1 - 2|F_2^m|^2\exp(-c_0 kt^2)$, for every $f_1, f_2 \in F_2^m$

$$\|f_1 - f_2\|_{L_2}^2 \le \|f_1 - f_2\|_{L_2^k}^2 + tL^2\|f_1 - f_2\|_{L_2}^2.$$

Thus, taking $t = 1/(2L^2)$, for $k \ge k_0(\delta, L)$ and $m \le c_8(L)k$, it is evident that with probability at least $1 - 2\exp(-c_9(L)k) \ge 1 - \delta/2$, for every $f_1, f_2 \in F_2^m$,

$$\|f_1 - f_2\|_{L_2}^2 \le 2\|f_1 - f_2\|_{L_2^k}^2,$$

as claimed. $\square$

4.3. *Shrinking properties of the $\gamma_{2,s}$ functionals.* The first corollary of Theorem 4.2 we shall present here is a shrinking property of $\gamma_{2,s}(F, L_2^I)$.

THEOREM 4.3. *For every $0 < \delta < 1$ there exists a constant $c(\delta) \sim \log(2/\delta)$ for which the following holds. Let F be an L-subgaussian class of functions on a probability space (Ω, μ) and assume that for every $f \in F$, $\mathbb{E}f = \alpha$ for some $\alpha \in \mathbb{R}$. Then, with probability at least $1 - \delta$, for every $I \subset \{1, \dots, k\}$ and every integer s that satisfies $2^s \le |I| \log(ek/|I|)$,*

$$\gamma_{2,s+1}(F, L_2^I) \le c(\delta) L \gamma_{2,s}(F, L_2) \sqrt{\log(ek/|I|)}.$$

Before proving Theorem 4.3, recall the following well-known result on the expectation of a monotone rearrangement of independent standard Gaussian variables (see, e.g., [5, 6]).

LEMMA 4.4. *Let $(g_i)_{i=1}^n$ be independent standard Gaussian variables and denote by $(g_i^*)_{i=1}^n$ the nonincreasing rearrangement of $(|g_i|)_{i=1}^n$. Then,*

$$\mathbb{E}g_i^* \sim \begin{cases} \sqrt{\log(2n/i)}, & \text{if } i \le n/2, \\ 1 - \dfrac{i}{n+1}, & \text{if } i > n/2. \end{cases}$$

Moreover,

$$\left(\mathbb{E} \sum_{i=1}^m (g_i^*)^2 \right)^{1/2} \sim \sqrt{m \log(en/m)}.$$

PROOF OF THEOREM 4.3. Fix $0 < \delta < 1$ and let the sets $\mathcal{A}, \mathcal{B} \in \Omega^k$ be as in the proof of Theorem 4.2. Take any $(X_1, \dots, X_k) \in \mathcal{A} \cap \mathcal{B}$, let $I \subset \{1, \dots, k\}$ and set $m = |I|$. Since $P_I F \subset P_I F_1^m + P_I F_2^m$, then by the sub-additivity of $\gamma_{2,s}$, it is evident that for every integer s,

$$\gamma_{2,s+1}(F, L_2^I) \le \gamma_{2,s}(F_1^m, L_2^I) + \gamma_{2,s}(F_2^m, L_2^I).$$

By (4.3), the mapping $P_I : (F_2^m, L_2) \to (F_2^m, L_2^I)$ is a Lipschitz function with a constant $c(\delta) L (\log(ek/m))^{1/2}$. Therefore, recalling that $F_2^m \subset F$,

$$\begin{aligned} \gamma_{2,s}(F_2^m, L_2^I) &\le c_1 \sqrt{\log(ek/m)} \gamma_{2,s}(F_2^m, L_2) \\ &\le c_1 \sqrt{\log(ek/m)} \gamma_{2,s}(F, L_2), \end{aligned} \tag{4.4}$$

where $c_1 = c_1(L, \delta) \sim L \log(2/\delta)$. To conclude the proof, observe that by Theorem 4.2, $P_I F_1^m \subset B_m W_m$, where $B_m = c_2(L, \delta) \gamma_{2,\tau_m}(F, L_2)$ and $W_m = \{x \in \mathbb{R}^m : x_j^* \le 1/\sqrt{j}, j = 1, \dots, m\}$.

Since the $\gamma_{2,s}$ functionals are monotone with respect to inclusion and are decreasing in s, and since $\|x\|_{L_2^I} = |I|^{-1/2}|x|$ for every $x \in \mathbb{R}^m$ then

$$\gamma_{2,s}(F_1^m, L_2^I) \le \gamma_2(F_1^m, L_2^I) \le B_m \frac{\gamma_2(W_m, |\cdot|)}{\sqrt{m}}.$$

Applying the Majorizing Measures theorem and Lemma 4.4

$$\gamma_2(W, |\cdot|) \le c_3 \mathbb{E} \sup_{w \in W} \sum_{i=1}^m g_i w_i = c_3 \mathbb{E} \sum_{i=1}^m \frac{g_i^*}{\sqrt{i}} \le c_4 \sqrt{m}.$$

Hence, for every s, $\gamma_{2,s}(F_1^m, L_2^I) \le c_4 B_m$, implying that for every $s \le \tau_m$, $\gamma_{2,s}(F_1^m, L_2^I) \le c_5(L,\delta)\gamma_{2,s}(F, L_2)$. Combining this with (4.4), it follows that for every $I \subset \{1, \dots, k\}$,

$$\gamma_{2,s+1}(F, L_2^I) \le c_6(L,\delta)\sqrt{\log(ek/|I|)}\gamma_{2,s}(F, L_2),$$

as claimed. □

REMARK 4.5. The proof of Theorem 4.3 yields a stronger result than the one formulated. It shows that with probability $1-\delta$, for every $I \subset \{1, \dots, k\}$ and every $s \ge 0$,

$$\gamma_{2,s+1}(F, L_2^I) \le c(L,\delta)(\gamma_{2,\tau_{|I|}}(F, L_2) + \sqrt{\log(ek/|I|)}\gamma_{2,s}(F, L_2)).$$

Observe that in some sense, the range $s \le \tau_{|I|}$ [i.e., $2^s \lesssim |I| \log(ek/|I|)$] is the interesting range of s, since

$$\gamma_{2,s}(P_I^\sigma F, |\cdot|) \le \operatorname{diam}(P_I^\sigma F, |\cdot|)\gamma_{2,s}(B_2^{|I|}, |\cdot|),$$

which decreases exponentially in s for $2^s \ge c_1|I|$.

Another outcome of Theorem 4.2 was formulated as Corollary B in the Introduction.

COROLLARY 4.6. *For every $0 < \delta < 1$ and $L > 0$, there exist a constant $c(\delta, L)$ such that the following holds. Let μ be an isotropic, L-subgaussian measure on $\mathbb{R}^n$, set $(X_i)_{i=1}^k$ to be independent, distributed according to μ and consider the random operator $\Gamma = \sum_{i=1}^k \langle X_i, \cdot\rangle e_i$. If $T \subset \mathbb{R}^n$ and $V = k^{-1/2}\Gamma T$, then with μ^k-probability at least $1-\delta$, for every $I \subset \{1, \dots, k\}$,*

$$\mathbb{E} \sup_{v \in V} \left| \sum_{i \in I} g_i v_i \right| \le c(L,\delta) \sqrt{\frac{|I|}{k} \log(ek/|I|)}\, \mathbb{E} \sup_{t \in T} \left| \sum_{i=1}^n g_i t_i \right|,$$

where the expectation on both sides is with respect to the Gaussian variables.

The proof of Corollary 4.6 follows from Theorem 4.2 and the Majorizing Measures theorem.

Proof of Corollary 4.6. Since μ is an L-subgaussian measure on $\mathbb{R}^n$, each $t \in \mathbb{R}^n$ corresponds to a function $f_t(x) = \langle t, x\rangle$, $f_t : \mathbb{R}^n \to \mathbb{R}$, for which $\|f_t\|_{\psi_2} \le L|t|$. Let $F = \{f_t : t \in T\}$, set $\Omega = \mathbb{R}^n$ and put $\sigma = (X_1, \dots, X_k) \in \Omega^k$ for which the assertion of Theorem 4.2 holds.

Fix $I \subset \{1, \dots, k\}$ of cardinality m. Since F is a class of linear functionals, the decomposition of F given in Theorem 4.2 actually implies a decomposition of T which we denote by T_1^m and T_2^m. Thus, for every $t \in T$, $t = t^1 + t^2$, where $t^i \in T_i^m$ for $i = 1, 2$. Since $P_\sigma F = \{(f_t(X_i))_{i=1}^k : t \in T\} = \Gamma T$ then

$$\mathbb{E} \sup_{v \in V} \left|\sum_{i \in I} g_i v_i\right| = \frac{1}{\sqrt{k}} \mathbb{E} \sup_{t \in T} \left|\sum_{i \in I} g_i \langle t_i^1 + t_i^2, X_i\rangle\right|.$$

Clearly, for any $u, v \in T$,

$$\|f_u - f_v\|_{L_2^I} = m^{-1/2} \|(\langle X_i, u - v\rangle)\|_{\ell_2^I} \quad \text{and} \quad \|f_u - f_v\|_{L_2} = \|u - v\|_{\ell_2^n},$$

and by the shrinking property of T_2^m, for every $u, v \in T_2^m$,

$$\|(\langle u, X_i\rangle)_{i=1}^n - (\langle v, X_i\rangle)_{i=1}^n\|_{\ell_2^I} \le c_1(L, \delta)(m \log(ek/m))^{1/2} \|u - v\|_{\ell_2^n}.$$

Therefore, by Slepian's lemma (see, e.g., [8]) and since $T_2^m \subset T$,

$$\begin{aligned} \frac{1}{\sqrt{k}} \mathbb{E}_g \sup_{t \in T} \left|\sum_{i \in I} g_i \langle t_i^2, X_i\rangle\right| &\le c_1(L, \delta) \sqrt{\frac{m}{k} \log(ek/m)}\, \mathbb{E}_g \sup_{t \in T} \left|\sum_{i=1}^n g_i t_i^2\right| \\ &\le c_1(L, \delta) \sqrt{\frac{m}{k} \log(ek/m)}\, \mathbb{E}_g \sup_{t \in T} \left|\sum_{i=1}^n g_i t_i\right|. \end{aligned}$$

Also, recall that $P_I^\sigma F_1^m \subset c_2(L, \delta) \gamma_{2,\tau_m}(F, L_2) W_m$, and, just as in the proof of Theorem 4.3 and by the isotropicity of μ,

$$\gamma_2(P_I^\sigma F_1^m, |\cdot|) \le c_3 \gamma_{2,\tau_m}(F, L_2)\sqrt{m} \le c_3 \gamma_2(F, L_2)\sqrt{m} = c_3 \gamma_2(T, |\cdot|)\sqrt{m}.$$

Applying the Majorizing Measures theorem,

$$\gamma_2(T, |\cdot|) \le c_4 \mathbb{E} \sup_{t \in T} \left|\sum_{i=1}^n g_i t_i\right|,$$

and thus

$$\begin{aligned} \frac{1}{\sqrt{k}} \mathbb{E} \sup_{t \in T} \left|\sum_{i \in I} g_i \langle t_i^1, X_i\rangle\right| &\le \frac{c_4}{\sqrt{k}} \gamma_2(P_I^\sigma F_1^m, |\cdot|) \\ &\le c_5 \sqrt{\frac{m}{k}}\, \mathbb{E} \sup_{t \in T} \left|\sum_{i=1}^n g_i t_i\right|, \end{aligned}$$

as claimed. □

To put Corollary 4.6 in the right context, even if one considers the case where μ is the canonical Gaussian measure on $\mathbb{R}^n$, the standard concentration estimate for the norm of a Gaussian vector around its mean (used in [14] to prove the result for $I = \{1, \ldots, k\}$) is not strong enough to allow a uniform control over all subsets of $\{1, \ldots, k\}$. What allows one to bypass this obstacle and obtain a result even in a subgaussian setup (in which case such a concentration result does not exist, and thus, even the result for $I = \{1, \ldots, k\}$ is not obvious) is the application of a cardinality-sensitive deviation argument rather than a concentration based method.

Note that the logarithmic term in Corollary 4.6 cannot be removed. For example, if $T = \{t\}$ and μ is the canonical Gaussian measure on $\mathbb{R}^n$ then the vector $(\langle X_i, t\rangle)_{i=1}^k$ has the same distribution as $|t|(\bar{g}_i)_{i=1}^k$, where $(\bar{g}_i)_{i=1}^k$ are independent standard Gaussian variables [that are also independent of $(g_i)_{i=1}^n$]. Recall that E_m is the collection of subsets of $\{1, \ldots, k\}$ of cardinality m and observe that

$$\begin{aligned}\mathbb{E}_X \sup_{I \in E_m} \mathbb{E}_g \left|\sum_{i \in I} g_i \langle X_i, t\rangle\right| &\sim \mathbb{E}_X \sup_{I \in E_m} \left(\sum_{i \in I} \langle X_i, t\rangle^2\right)^{1/2} \\ &= |t|\mathbb{E} \sup_{I \in E_m} \left(\sum_{i \in I} (\bar{g}_i)^2\right)^{1/2} = |t|\mathbb{E}\left(\sum_{i=1}^m (\bar{g}_i^*)^2\right)^{1/2} \\ &\sim |t|\sqrt{m \log(ek/m)},\end{aligned}$$

where the last assertion is the second part of Lemma 4.4. Therefore, with probability at least c_1, there will be some $I \in E_m$ for which

$$\frac{1}{\sqrt{k}} \mathbb{E}_g \left|\sum_{i \in I} g_i \langle X_i, t\rangle\right| \geq c_2 |t| \sqrt{\frac{m}{k} \log(ek/m)},$$

showing that indeed, one cannot remove the logarithmic term.

5. Proof of Theorem A. As we explained in previous sections, our method of selecting signs in a way that is better than choosing typical signs depends on two properties. One is that the complexity of the set (as captured, e.g., by $\gamma_{2,s}$ or the metric entropy of the set) is small, and the other is that the set is in a good position (e.g., if coordinate projections shrink the set's complexity). Our results thus far indicate that for a subgaussian class and a typical $\sigma = (X_i)_{i=1}^k$, $P_\sigma F$ is essentially a set in a good position. Thus, it seems likely that the ability to choose signs that outperform the typical behavior of signs will be governed solely by the complexity of F. As Theorem A, which we reformulate below, shows, this is indeed the case.

Although the proof of Theorem A is rather technical, the basic idea behind it is simple. It follows from a combination of the two main results of the previous sections. First of all, that a typical coordinate projection of a subgaussian class is contained in the Minkowski sum of a small set and a set that satisfies a strong shrinking property. Second, that the discrepancy of sets that satisfy a shrinking property may be bounded in a nontrivial manner using their metric complexity.

THEOREM 5.1. *For any $0<\delta<1$, $0<\rho<1/2$ and $L>0$ there are constants c_1 and c_2 that depend on δ, ρ and L and for which the following holds. Let $F \subset L_2(\mu)$ be an L-subgaussian class, consisting of mean zero functions. Then, for every k there is a set $\mathcal{A}_k \subset \Omega^k$ with $\mu^k(\mathcal{A}_k) \geq 1-\delta$ such that for every $(X_1,\ldots,X_k) \in \mathcal{A}_k$ and every $I \subset \{1,\ldots,k\}$,*

$$\inf_{(\varepsilon_i)_{i\in I}} \sup_{f\in F} \left| \sum_{i\in I} \varepsilon_i f(X_i) \right| \leq \sqrt{|I|}\, a_{|I|},$$

where for every $n \leq k$

$$a_n \leq c_1 \left(\gamma_{2,\log_2\log_2(c_2 n)}(F,L_2) \cdot \sqrt{\log(ek/n)} + \operatorname{diam}(F,L_2) \frac{\log k}{n^{1/2-\rho}} \right).$$

Before proving Theorem 5.1, let us recall the following notation. For every integer m, s_m is the largest integer s such that $2^{2^{s+1}} \leq \kappa_3 m$. If $m \leq k$, then τ_m is the first integer for which $2^{2^s} \geq \exp(m \cdot \log(ek/m))$. In particular, for every $1 \leq m \leq k$, $\tau_m \geq \log_2 \log_2 k$ (but of course, τ_m could be much larger). We will also say that for $\sigma = (X_1,\ldots,X_k)$, a function class F satisfies the shrinking property on $I \subset \{1,\ldots,k\}$ with a constant c if for every $f,h \in F$,

$$\|f-h\|_{L_2^I} \leq c\sqrt{\log(ek/|I|)}\, \|f-h\|_{L_2}.$$

PROOF OF THEOREM 5.1. Fix $0<\delta<1$ and consider $\sigma=(X_1,\ldots,X_k)$ for which the assertions of Theorem 4.2 hold. Fix any integer $n \leq k$ and let $I_0 \subset \{1,\ldots,k\}$ be of cardinality n. Using the notation of Theorem 4.2, we may decompose $F \subset F_1^n + F_2^n$, where $P_{I_0}^\sigma F_1^n \subset c_1 \gamma_{2,\tau_n}(F,L_2) W_n$, $F_2^n \subset F$, and F_2^n satisfies the shrinking property on every $I \subset \{1,\ldots,k\}$ of cardinality n with a constant $c=c(L,\delta)$—and in particular, it does so on I_0.

For every $f \in F$, choose $f_1 \in F_1^n$ and $f_2 \in F_2^n$ such that $f = f_1 + f_2$. Hence, for every $(\eta_i)_{i\in I_0} \in \{-1,0,1\}^{I_0}$ and any $f \in F$,

$$\left| \sum_{i\in I_0} \eta_i f(X_i) \right| \leq c_1 \gamma_{2,\tau_n}(F,L_2)\sqrt{n} + \left| \sum_{i\in I_0} \eta_i f_2(X_i) \right|.$$

Let $(F_s)_{s=1}^{\infty}$ be an admissible sequence of F_2^n which will be specified later and set $\pi_s(f_2)$ to be a nearest point to f_2 in F_s. As in Corollary 3.4, if

$$Q_s^0 = \kappa_4 \begin{cases} \exp(-\kappa_5 n^{1/2}), & \text{if } s < s_n, \\ 1, & \text{if } s = s_n, \\ 2^{s/2}, & \text{if } s > s_n, \end{cases}$$

then there exist $(\eta_i^0)_{i \in I_0} \in \{-1,0,1\}^{I_0}$ such that $n/4 \le |\{i : \eta_i^0 = 0\}| \le 3n/4$ and for every $(f_2(X_i))_{i \in I_0} \in P_{I_0}^{\sigma} F_2^n$,

$$\left| \sum_{i \in I_0} \eta_i^0 f_2(X_i) \right| \le \sum_{s=1}^{\infty} Q_s^0 \|\pi_s(f_2) - \pi_{s-1}(f_2)\|_{\ell_2^{I_0}}.$$

Since functions in F_2^n satisfy the shrinking property with a constant c, then for every $f \in F_2^n$

$$\|\pi_s(f) - \pi_{s-1}(f)\|_{\ell_2^{I_0}} \le c\sqrt{n \log(ek/n)} \|\pi_s(f) - \pi_{s-1}(f)\|_{L_2},$$

implying that

$$\left| \sum_{i \in I_0} \eta_i^0 f_2(X_i) \right| \le c\sqrt{n \log(ek/n)} \sum_{s=1}^{\infty} Q_s^0 \|\pi_s(f_2) - \pi_{s-1}(f_2)\|_{L_2}.$$

Let $I_1 = \{i \in I_0 : \eta_i = 0\}$ and continue in the same manner: first decompose $F \subset F_1^{|I_1|} + F_2^{|I_1|}$, then apply the fact that $P_{I_1}^{\sigma} F_1^{|I_1|}$ is contained in an appropriate weak ℓ_2 ball, and finally, since $F_2^{|I_1|}$ satisfies the shrinking property on I_1, use Corollary 3.4 again, and so on.

As a result of iterating this argument, there are nested subsets of $\{1, \dots, k\}$, $(I_j)_{j=0}^{j_0}$, with $|I_0| = n$ and of cardinalities

$$\frac{|I_j|}{4} \le |I_{j+1}| \le \frac{3}{4}|I_j|, \qquad 1 \le |I_{j_0}| \le 10,$$

and vectors $(\eta_i^j)_{i \in I_j} \in \{-1,0,1\}^{I_j}$, such that $I_{j+1} = \{i : \eta_i^j = 0\}$ with the following property. For every $0 \le j \le j_0$, let

$$Q_s^j = \kappa_4 \begin{cases} \exp(-\kappa_5 |I_j|^{1/2}), & \text{if } s < s_{|I_j|}, \\ 1, & \text{if } s = s_{|I_j|}, \\ 2^{s/2}, & \text{if } s > s_{|I_j|}, \end{cases}$$

and for every $f \in F$, $f = f_1^j + f_2^j$, $f_1^j \in F_1^{|I_j|}$, $f_2^j \in F_2^{|I_j|}$ one has

$$\left| \sum_{i \in I_j} \eta_i^j f(X_j) \right| \le \sum_{s=1}^{\infty} Q_s^j \|\pi_s(f_2^j) - \pi_{s-1}(f_2^j)\|_{\ell_2^{I_j}} + c_1 \sqrt{|I_j|} \gamma_{2,\tau_{|I_j|}}(F, L_2)$$

$$\leq c\sqrt{|I_j|\log(ek/|I_j|)}\sum_{s=1}^{\infty} Q_s^j\|\pi_s(f_2^j)-\pi_{s-1}(f_2^j)\|_{L_2}$$
$$+ c_1\sqrt{|I_j|}\gamma_{2,\tau_{|I_j|}}(F,L_2).$$

Therefore, there are signs $(\varepsilon_i)_{i=1}^n \in \{-1,1\}^n$ such that,

$$\begin{aligned}
&\sup_{f\in F}\left|\sum_{i=1}^n \varepsilon_i f(X_i)\right| \\
&\leq c\sum_{j=0}^{j_0}\sqrt{|I_j|\log(ek/|I_j|)}\sup_{f\in F}\sum_{s=1}^{\infty} Q_s^j\|\pi_s(f)-\pi_{s-1}(f)\|_{L_2} \qquad (5.1)\\
&\quad + c_1\sum_{j=0}^{j_0}\sqrt{|I_j|}\gamma_{2,\tau_{|I_j|}}(F,L_2) + c_2\,\mathrm{diam}(F,L_2)\log(ek),
\end{aligned}$$

where the last term comes from a trivial estimate on the discrepancy of a projection of F onto the set of coordinates $\{i\in I_{j_0}: \eta_i^{j_0}=0\}$ and the shrinking phenomenon.

To complete the proof, one has to bound (5.1) from above. To that end, set $b_j=|I_j|$ and recall that $(1/4)^j n \leq b_j \leq (3/4)^j n$. To estimate the second term in (5.1), since $b_j\geq 1$ then $\tau_{b_j}\geq \log_2\log_2 k$. Therefore,

$$\sum_{j=0}^{j_0}\sqrt{|I_j|}\gamma_{2,\tau_{|I_j|}}(F,L_2)\leq c_3\sqrt{n}\gamma_{2,\log_2\log_2 k}(F,L_2).$$

Turning our attention to the first term in (5.1), for every $1\leq \ell\leq s_n$ let $U_\ell=\{j: s_{b_j}=\ell\}$, set $b_\ell^+=\max\{b_j: j\in U_\ell\}$ and $b_\ell^-=\min\{b_j: j\in U_\ell\}$. In other words, U_ℓ consists of all the integers j for which $s_{|I_j|}=s_{b_j}=\ell$; b_ℓ^+ is the largest cardinality of such a set and b_ℓ^- is the smallest one. Since

$$\kappa_3^{-1}2^{2^{\ell+1}}\leq b_\ell^-\leq b_\ell^+\leq \min\{\kappa_3^{-1}2^{2^{\ell+2}},n\},$$

then for every $j\in U_\ell$, the sequence $(Q_s^j)_{s>0}$ satisfies

$$Q_s^j\leq \kappa_4\begin{cases}\exp(-\kappa_5\kappa_3^{-1/2}\cdot 2^{2^\ell}), & \text{if } s<\ell,\\ 1, & \text{if } s=\ell,\\ 2^{s/2}, & \text{if } s>\ell,\end{cases}$$

and we denote this sequence $(Q_s^\ell)_{s>0}$. Since b_j decays exponentially, then

$$\sum_{j=0}^{j_0}\sqrt{|I_j|\log(ek/|I_j|)}\sup_{f\in F}\sum_{s=1}^{\infty} Q_s^j\|\pi_s(f)-\pi_{s-1}(f)\|_{L_2}$$

$$
\begin{aligned}
&= \sum_{\ell=1}^{s_n} \sum_{b_j \in U_\ell} \sqrt{b_j \log(ek/b_j)} \sup_{f\in F} \sum_{s=1}^{\infty} Q_s^j \|\pi_s(f) - \pi_{s-1}(f)\|_{L_2} \\
&\le c_4 \sum_{\ell=1}^{s_n} \sqrt{b_\ell^+ \log(ek/b_\ell^+)} \sup_{f\in F} \sum_{s=1}^{\infty} Q_s^\ell \|\pi_s(f) - \pi_{s-1}(f)\|_{L_2}.
\end{aligned}
$$

Set $d_\ell = \sqrt{b_\ell^+ \log(ek/b_\ell^+)}$, fix $0<\rho<1/2$ and let ℓ_1 be the largest integer such that $\kappa_3^{-1} 2^{2^{\ell_1+2}} \le n^{2\rho}$. Then, for every $\ell \le \ell_1$, $b_\ell^+ \le \kappa_3^{-1} 2^{2^{\ell_1+2}} \le n^{2\rho}$ and for $\ell > \ell_1$, $b_\ell^+ \le n$. Observe that for every s, ℓ, $Q_s^\ell \le \kappa_4 2^{s/2}$ and for every $f \in F$, $\|\pi_s(f) - \pi_{s-1}(f)\|_{L_2} \le 2\,\mathrm{diam}(F, L_2)$. Therefore,

$$
\begin{aligned}
&\sum_{\ell=1}^{s_n} d_\ell \cdot \sup_{f\in F} \sum_{s=1}^{\infty} Q_s^\ell \|\pi_s(f) - \pi_{s-1}(f)\|_{L_2} \\
&\qquad \le \sum_{\ell=1}^{s_n} d_\ell \cdot \sup_{f\in F} \sum_{s=1}^{\ell_1} Q_s^\ell \|\pi_s(f) - \pi_{s-1}(f)\|_{L_2} \\
&\qquad\quad + \sum_{\ell=1}^{s_n} d_\ell \cdot \sup_{f\in F} \sum_{s=\ell_1+1}^{\infty} Q_s^\ell \|\pi_s(f) - \pi_{s-1}(f)\|_{L_2} \\
&\qquad \le 2\,\mathrm{diam}(F, L_2) \sum_{\ell=1}^{s_n} \sum_{s=1}^{\ell_1} d_\ell Q_s^\ell \\
&\qquad\quad + \kappa_4 \sum_{\ell=1}^{s_n} d_\ell \cdot \sup_{f\in F} \sum_{s=\ell_1+1}^{\infty} 2^{s/2} \|\pi_s(f) - \pi_{s-1}(f)\|_{L_2} \\
&\qquad \le 2\,\mathrm{diam}(F, L_2) \sum_{\ell=1}^{s_n} \sum_{s=1}^{\ell_1} d_\ell Q_s^\ell + c_5(\rho) \sqrt{n \log(ek/n)} \cdot \gamma_{2,\ell_1}(F, L_2)
\end{aligned}
$$

for an almost optimal choice of $(F_s)_{s=\ell_1}^{\infty}$.

Now, for every $s \le \ell_1$ and using that $b_\ell^+ \le n^{2\rho}$ for $\ell < \ell_1$ and $b_\ell^+ \le n$ for $\ell \ge \ell_1$, it is evident that

$$
\begin{aligned}
\sum_{\ell=1}^{s_n} d_\ell Q_s^\ell &= \sum_{\ell=1}^{s_n} \sqrt{b_\ell^+ \log(ek/b_\ell^+)} Q_s^\ell \\
&\le n^\rho \sqrt{\log(ek/n^{2\rho})} \sum_{\ell<\ell_1} Q_s^\ell + \sqrt{n \log(ek/n)} \sum_{\ell \ge \ell_1} Q_s^\ell = (*).
\end{aligned}
$$

Note that $2^{\ell_1+3} \geq 2\rho \log_2 c_6 n$ and thus $2^{\ell_1+1} \geq (\rho/2)\log_2 c_6 n$. Therefore, there is an absolute constant c_7 such that if $s \leq \ell_1$ then

$$\sum_{\ell \leq \ell_1} Q_s^\ell = \sum_{\ell \leq s} Q_s^\ell + \sum_{\ell=s+1}^{\ell_1} Q_s^\ell \\ \leq c_7(s2^{s/2} + \exp(-c_3 2^{2^s})) \leq 2c_7 s 2^{s/2}$$

and

$$\sum_{\ell > \ell_1} Q_s^\ell \leq c_7 \exp(-c_3 2^{2^{\ell_1}}) \leq c_7 \exp(-c_8 n^{\rho/2}).$$

Hence, there is a constant $c_9(\rho)$ such that for every $s \leq \ell_1$,

$$(*) \leq c_9(\rho) s 2^{s/2} n^\rho \sqrt{\log(ek/n^{2\rho})},$$

and thus,

$$\begin{aligned}\operatorname{disc}(P_I^\sigma F) \leq{}& c_2 \operatorname{diam}(F, L_2)\log(ek) + c_3\sqrt{n}\gamma_{2,\log_2\log_2 k}(F, L_2) \\ &+ c_{10}\operatorname{diam}(F, L_2)\cdot \ell_1 2^{\ell_1/2} n^\rho \sqrt{\log(ek/n^{2\rho})} \\ &+ c_{10}\gamma_{2,\ell_1}(F, L_2)\cdot\sqrt{n\log(ek/n)}.\end{aligned}$$

Since $(\rho/4)\log_2 c_6 n \leq 2^{\ell_1} \leq (\rho/2)\log_2 c_6 n$, the claim follows. $\square$

Corollary 5.2. *Let $0 < \rho < 1/2$. Under the assumptions of Theorem 5.1 and using its notation, for every $\sigma \in \mathcal{A}_k$*

$$\operatorname{disc}(P_\sigma F) \leq c_1(\sqrt{k}\cdot\gamma_{2,\log_2\log_2(c_2 k)}(F, L_2) + k^\rho \operatorname{diam}(F, L_2))$$

and

$$\operatorname{Hdisc}(P_\sigma F) = \sup_{I\subset\{1,\ldots,k\}} \inf_{(\varepsilon_i)_{i=1}^k} \sup_{f\in F}\left|\sum_{i\in I}\varepsilon_i f(X_i)\right| \leq \sup_{1\leq n\leq k} a_n\sqrt{n}.$$

In particular, if $\lim_{s\to\infty}\gamma_{2,s}(F, L_2) = 0$ (i.e., if F is μ-pregaussian), then

$$\frac{1}{\sqrt{k}}\operatorname{Hdisc}(\{(f(X_i))_{i=1}^k : f \in F\})$$

converges in probability to 0.

Let us mention once again that the reason that Theorem 5.1 is meaningful is because for a typical $(X_i)_{i=1}^k$, a class of mean zero functions that is L-subgaussian satisfies that

$$c_1\sigma_F\sqrt{k} \leq \mathbb{E}_\varepsilon \sup_{f\in F}\left|\sum_{i=1}^k \varepsilon_i f(X_i)\right| \leq c_2(L)\gamma_2(F, L_2)\sqrt{k}.$$

Thus, there is a true gap between the discrepancy (or even the hereditary discrepancy) of a typical coordinate projection and the average over signs of a coordinate projection of a pregaussian, subgaussian class F.

6. Equivalence for large sets. In this section, our aim is to show that if F is a subgaussian class that indexes a bounded Gaussian process, then the reason for the gap between the expectation over signs of a random coordinate projection and the infimum over signs is indeed that $\lim_{s\to\infty}\gamma_{2,s}(F,L_2)=0$.

To be more precise, we show the following.

THEOREM 6.1. *For every $0<\delta<1$ and $A,B,L>0$ there is a constant $c(\delta,A,B,L)$ for which the following holds. Let $F\subset B(L_2(\mu))$ be a class of mean zero functions such that* $\operatorname{absconv}(F)$ *is L-subgaussian. If $\gamma_2(F,L_2(\mu))\le A<\infty$ and if the entropy numbers satisfy that*

$$\limsup_{j\to\infty} j^{1/2}e_j(\operatorname{absconv}(F),L_2(\mu))=B>0,$$

then there is a sequence of integers $(k_i)_{i=1}^{\infty}$ tending to infinity, such that for every i, with probability at least $1-\delta$ in Ω^{k_i},

$$\operatorname{Hdisc}(P_\sigma F)\ge c(\delta,A,B,L)\sqrt{k_i},$$

where $\sigma=(X_1,X_2,\dots,X_{k_i})\in\Omega^{k_i}$ is selected according to μ^{k_i}. In particular, $\operatorname{Hdisc}(P_\sigma F)/\sqrt{k}$ *does not converge to 0 in probability.*

Observe that this is almost the reverse direction of Theorem A. Indeed, it is well known (see, e.g., [3], Chapter 9) that there is no entropic characterization of classes that index a bounded Gaussian process which is not continuous; such a characterization is given by a majorizing measures argument [17]. However, because $\{G_f: f\in F\}$ is a bounded process with a covariance structure endowed by $L_2(\mu)$, then by Sudakov's inequality (see, e.g., [8], Chapter 3),

$$\log N(\varepsilon,\operatorname{absconv}(F),L_2(\mu))\le c_1\left(\frac{\mathbb{E}\sup_{f\in F}G_f}{\varepsilon}\right)^2.$$

On the other hand, since $F\subset B(L_2(\mu))$ is not μ-pregaussian, one can show that

$$\int_0^1\sqrt{\log N(\varepsilon,F,L_2(\mu))}\,d\varepsilon=\infty.$$

Thus, up to a logarithmic factor, the entropy numbers of F are as in Theorem 6.1. Whether Theorem 6.1 remains true using only the assumption that $\limsup_{s\to\infty}\gamma_{2,s}(F,L_2(\mu))>0$ is not clear.

The idea behind the proof of Theorem 6.1 is to find a cube in a typical coordinate projection of $\operatorname{absconv}(F)$. We will first show that if $\operatorname{absconv}(F)$ has a "large" separated set with respect to the $L_2(\mu)$ metric at scale $\sim 1/\sqrt{k}$, then its typical coordinate projection of dimension k contains a cubic structure of dimension $\sim k$ and scale $\sim 1/\sqrt{k}$. The cubic structure we will be interested in is captured by the combinatorial dimension.

DEFINITION 6.2. Let F be a class of functions on Ω. For every $\varepsilon > 0$, a set $\sigma = \{x_1, \dots, x_j\} \subset \Omega$ is said to be ε-shattered by F if there is some function $s : \sigma \to \mathbb{R}$, such that for every $I \subset \{1, \dots, j\}$ there is some $f_I \in F$ for which $f_I(x_i) \geq s(x_i) + \varepsilon$ if $i \in I$, and $f_I(x_i) \leq s(x_i) - \varepsilon$ if $i \notin I$. Define the combinatorial dimension at scale ε by

$$\mathrm{VC}(F, \varepsilon) = \sup\{|\sigma| | \sigma \subset \Omega, \sigma \text{ is } \varepsilon\text{-shattered by} F\}.$$

Note that if F is a $\{0,1\}$-class of functions then $\mathrm{VC}(F) = \mathrm{VC}(F, 1/2)$. Also, in a similar way one may define the combinatorial dimension of a subset of $\mathbb{R}^n$, when each vector is viewed as a function defined on $\{1, \dots, n\}$.

It is standard to verify that if $\mathrm{VC}(F, \varepsilon) \geq m$, then the coordinate projection $P_\tau F$, defined by the shattered set τ, contains a subset of cardinality $\exp(cm)$ which is $c_1\varepsilon$-separated with respect to the $L_2(\mu_\tau)$ norm (recall that μ_τ is the uniform probability measure supported on τ), and that $\mathrm{disc}(P_\tau F) \geq c_2 m\varepsilon$ (see Lemma 6.5). As we mentioned in the Introduction, the reverse direction is also true, and if $F \subset B(L_\infty(\Omega))$ contains a large well-separated set in $L_2(\mu)$ that it must have a large combinatorial dimension at a scale that is proportional to the scale of the separation (see [12] for an exact statement and proof). A fact that will be used here and which is based on this reverse direction is the following.

THEOREM 6.3 [12]. *There exist absolute constants c_1 and c_2 for which the following holds. Let $V \subset B_\infty^k$ and assume that $\mathbb{E}\sup_{v \in V} |\sum_{i=1}^k \varepsilon_i v_i| \geq \delta k$. Then,*

$$\mathrm{VC}(V, c_1\delta) \geq c_2 \delta^2 k.$$

Hence, the only reason that $\mathbb{E}\sup_{v \in V} |\sum_{i=1}^k \varepsilon_i v_i|$ is almost extremal is that V contains a large cube in a high-dimensional coordinate projection.

The key observation of this section is the following theorem.

THEOREM 6.4. *For every $A, B, L > 0$ and $0 < \delta < 1$ there exist constants c_1 and c_2 that depend on A, B, L and δ for which the following holds. Let $F \subset B(L_2(\mu))$ be a convex, symmetric, L-subgaussian set of mean zero functions. Suppose that $\gamma_2(F, \psi_2) \leq A < \infty$ and that there is some k for which $e_k(F, L_2(\mu)) \geq B/\sqrt{k}$. Then, there is a set $\Sigma \subset \Omega^k$ such that $\mu^k(\Sigma) \geq 1 - \delta$ and for every $\sigma \in \Sigma$,*

$$\mathrm{VC}\left(P_\sigma F, \frac{c_1}{\sqrt{k}}\right) \geq c_2 k.$$

Theorem 6.4 implies Theorem 6.1 because of the next lemma.

LEMMA 6.5. *If $T \subset \mathbb{R}^n$, then*

$$\mathrm{Hdisc}(T) \geq \sup_{\delta>0} \delta \, \mathrm{VC}(\mathrm{absconv}(T), \delta).$$

PROOF. First, note that $\mathrm{Hdisc}(T) = \mathrm{Hdisc}(\mathrm{absconv}(T))$, and thus we may assume that T is convex and symmetric. Now, let $I \subset \{1, \dots, n\}$ be δ-shattered by T with the level function s. Fix $(\varepsilon_i)_{i \in I} \in \{-1, 1\}^{|I|}$ and without loss of generality assume that $\sum_{i \in I} \varepsilon_i s_i \geq 0$. Since I is δ-shattered by T, there is some $t' \in T$ for which $t'_i \geq s_i + \delta$ when $\varepsilon_i = 1$ and $t'_i \leq s_i - \delta$ when $\varepsilon_i = -1$. Thus,

$$\sup_{t \in T} \left| \sum_{i \in I} \varepsilon_i t_i \right| \geq \left| \sum_{i \in I} \varepsilon_i (t'_i - s_i) + \sum_{i \in I} \varepsilon_i s_i \right| \geq \left| \sum_{i \in I} \varepsilon_i (t'_i - s_i) \right| \geq |I| \delta,$$

as claimed. $\square$

Hence, from here on we may assume without loss of generality that the class F is convex and symmetric, and that it is L-subgaussian.

The proof of Theorem 6.4 requires several additional facts. To formulate them, denote for $V \subset \mathbb{R}^n$

$$\ell_*(V) = \mathbb{E} \sup_{v \in V} \left| \sum_{i=1}^n g_i v_i \right|,$$

and if $A, B \subset \mathbb{R}^n$, set $N(A, B)$ to be the minimal number of translates of B needed to cover A.

The first lemma we need is taken from [9].

LEMMA 6.6. *Let $V \subset \mathbb{R}^k$ be a convex, symmetric set. For $\rho > 0$, set $V_\rho = V \cap \rho B_2^k$ and $F(\rho) = \ell_*(V)/\ell_*(V_\rho)$. Then,*

$$N(V, 8\rho B_2^k) \leq \exp\left(2 \left(\frac{\ell_*(V_\rho)}{\rho} \right)^2 \log(6F(\rho)) \right).$$

The second result was proved in [13] (Theorem 2.3). Although it was formulated there for subsets of $\mathbb{R}^n$, its proof shows that the claim is true for any subgaussian class of functions. It implies that a random coordinate projection of F, viewed as a mapping between $L_2(\mu)$ and L_2^k, is almost norm preserving for functions with a sufficiently large $L_2(\mu)$ norm.

THEOREM 6.7. *There exist absolute constants c_1 and c_2 for which the following holds. Let $F \subset L_2(\mu)$ be a convex, symmetric, L-subgaussian class of functions. For every $\theta > 0$ and any positive integer k, set*

$$r_k(\theta) = \inf\left\{ \rho : \rho \geq \frac{\gamma_2(F \cap \rho B(L_2(\mu)), \psi_2)}{\theta \sqrt{k}} \right\}.$$

Then, with probability at least $1-2\exp(-c_1\theta^2 k/L^4)$*, for every* $f\in F$ *such that* $\|f\|_{L_2(\mu)}\geq r_k(\theta/c_2L^2)$*,*

$$(1-\theta)^{1/2}\|f\|_{L_2(\mu)}\leq \|f\|_{L_2^k}\leq (1+\theta)^{1/2}\|f\|_{L_2(\mu)}.$$

COROLLARY 6.8. *For every* $L>0$*, there are constants* κ_6 *and* κ_7 *that depend only on* L*, for which the following holds. Let* F *be an* L*-subgaussian, convex and symmetric class of functions for which* $\gamma_2(F,\psi_2)\leq A<\infty$*. Then, with probability at least* $1-2\exp(-\kappa_6 k)$*, if* $f\in F$ *and* $\|f\|_{L_2(\mu)}\geq \kappa_7 A/\sqrt{k}$ *then*

$$\sqrt{\tfrac{1}{2}}\|f\|_{L_2(\mu)}\leq \|f\|_{L_2^k}\leq \sqrt{\tfrac{3}{2}}\|f\|_{L_2(\mu)}.$$

In particular, if $H\subset F$ *is an* ε*-separated set in* $L_2(\mu)$ *for* $\varepsilon>2\kappa_7 A/\sqrt{k}$ *then with probability at least* $1-2\exp(-\kappa_6 k)$*,* $P_\sigma H$ *is* $\varepsilon/4$*-separated in* L_2^k*.*

PROOF. Let c_1 and c_2 be as in Theorem 6.7. Observe that

$$\gamma_2(F\cap \rho B(L_2(\mu)),\psi_2)\leq \gamma_2(F,\psi_2)\leq A$$

and apply Theorem 6.7 for $\theta=1/2$. Thus, $r_k(\theta/c_2L)\leq c_3(L)A/\sqrt{k}$, implying that if $c_4(L)=c_1/4L^4$ then with probability at least $1-2\exp(-c_4(L)k)$, if $\|f\|_{L_2(\mu)}\geq c_3(L)A/\sqrt{k}$ then

$$\tfrac{1}{2}\|f\|_{L_2(\mu)}^2\leq \|f\|_{L_2^k}^2\leq \tfrac{3}{2}\|f\|_{L_2(\mu)}^2.$$

Turning to the second part, note that if $H\subset F$ is ε-separated in $L_2(\mu)$ for $\varepsilon>2c_1(L)A/\sqrt{k}$, then for every $h_1,h_2\in H$, $f=(h_1-h_2)/2\in F$ and $\|f\|_{L_2(\mu)}\geq c_1(L)A/\sqrt{k}$. Thus, the second part follows from the first one. □

Now we can formulate the first localization result, showing that the richness of a typical coordinate projection comes from the intersection of F with a ball of radius $\sim 1/\sqrt{k}$.

THEOREM 6.9. *For every positive* A*,* B*,* L *and* $0<\delta<1$*, there are constants* $c>1$*,* c_1 c_2 *and* c_3 *depending on* A*,* B*,* L *and* δ *for which the following holds. Let* $F\subset B(L_2(\mu))$ *be a convex, symmetric,* L*-subgaussian class of mean zero functions such that* $\gamma_2(F,\psi_2)\leq A<\infty$*. Fix an integer* k *and assume that* $e_k(F,L_2(\mu))\geq B/\sqrt{k}$*. Then, with probability at least* $1-\delta-2\exp(-c_1k)$*,*

$$\mathbb{E}_g \sup_{f\in F\cap c_2/\sqrt{k}B(L_2(\mu))}\left|\sum_{i=1}^{ck} g_i f(X_i)\right|\geq c_3\sqrt{k}.$$

PROOF. Since F is L-subgaussian and by applying Sudakov's inequality, we may assume without loss of generality that $A/B > 1$. Let H be a maximal $B/\sqrt{k}$ separated set in F with $\log|H| \geq k$. Let $k' = c^2 k$ for a constant $c > 1$ to be named later. Since H is $\varepsilon = cB/\sqrt{k'}$ separated in $L_2(\mu)$, then by Corollary 6.8, with probability at least $1 - 2\exp(-\kappa_6 k') = 1 - 2\exp(-c_1(L)k)$, if $\varepsilon \geq 2\kappa_7 A/\sqrt{k'}$ then $P_\sigma H$ is $\varepsilon/4$-separated in $L_2^{k'}$. Moreover, if $f \in F$ satisfies $\|f\|_{L_2(\mu)} \geq \kappa_7 A/\sqrt{k'}$ then

$$\tfrac{1}{2}\|f\|^2_{L_2(\mu)} \leq \|f\|^2_{L_2^{k'}} \leq \tfrac{3}{2}\|f\|^2_{L_2(\mu)}. \tag{6.1}$$

Clearly, the condition on ε holds if $c \sim_L A/B$, and since $c > 1$ it follows that $k' > k$.

Consider the set $U = \operatorname{absconv}(H)$. By the Majorizing Measures theorem and a simple application of Theorem 4.2, with probability at least $1 - \delta$ for $|\sigma| = k'$,

$$\ell_*(P_\sigma U) \leq c_2 \gamma_2(P_\sigma U, |\cdot|) \leq c_3(L,\delta) A\sqrt{k'}. \tag{6.2}$$

Let $\sigma = (X_i)_{i=1}^{k'}$ be in the intersection of the two events given by (6.1) and (6.2), set $V = P_\sigma U$ and note that $B(L_2^\sigma) = \sqrt{k'} B_2^{k'}$. Therefore,

$$\begin{aligned} k &\leq \log N\left(V, \frac{\varepsilon}{4}\sqrt{k'} B_2^{k'}\right) = \log N(V, c_4 B_2^{k'}) \\ &\equiv \log N(V, 8\rho B_2^{k'}), \end{aligned} \tag{6.3}$$

where $c_4 \sim_L A$ (and thus $\rho \sim_L A$ as well). If $V_\rho = V \cap \rho B_2^{k'}$, then by Lemma 6.6, (6.2) and (6.3),

$$k \leq 2\left(\frac{\ell_*(V_\rho)}{\rho}\right)^2 \log(6F(\rho)) \leq 2\left(\frac{\ell_*(V_\rho)}{\rho}\right)^2 \log\left(\frac{c_3 A\sqrt{k'}}{\ell_*(V_\rho)}\right).$$

Solving this inequality for $\ell_*(V_\rho)$, it is evident that there exists a constant $c_5 \sim_{L,\delta} B/\sqrt{\log(c_6 A^2/B^2)}$ (where c_6 depends on L and δ) for which $\ell_*(V_\rho) \geq c_5\sqrt{k'}$. Since F is convex and symmetric and $H \subset F$, then

$$V_\rho = P_\sigma(\operatorname{absconv}(H)) \cap \rho B_2^{k'} \subset P_\sigma\left(\left\{f \in F : \|f\|_{L_2^{k'}} \leq \frac{\rho}{\sqrt{k'}}\right\}\right)$$

and by (6.1),

$$\left\{f \in F : \|f\|_{L_2^{k'}} \leq \frac{\rho}{\sqrt{k'}}\right\} \subset \left\{f \in F : \|f\|_{L_2(\mu)} \leq 2\frac{\max\{\rho, \kappa_7 A\}}{\sqrt{k'}}\right\}.$$

Hence, there is a constant $c_7 \sim_{L,\delta} A$ for which with probability at least $1 - \delta - 2\exp(-c_1 k)$,

$$V_\rho \subset P_\sigma\left(\left\{f \in F : \|f\|_{L_2(\mu)} \leq \frac{c_7}{\sqrt{k'}}\right\}\right),$$

implying that

$$c_5\sqrt{k'} \le \mathbb{E}_g \sup_{\{f\in F:\, \|f\|_{L_2(\mu)}\le c_7/\sqrt{k'}\}} \left|\sum_{i=1}^{k'} g_i f(X_i)\right|. \qquad \square$$

The next step in the proof of Theorem 6.4 is a second localization argument. Theorem 6.9 shows that under our assumptions, there is a small ball (of radius $\sim 1/\sqrt{k}$) in F that causes coordinate projections of F of dimension k to be "rich." Now, one has to localize even further by truncating the functions in $F_1 = F\cap (c/\sqrt{k})B(L_2(\mu))$.

Definition 6.10. For every $\beta > 0$ and every $f\in F$, let

$$f_\beta^- = f\mathbb{1}_{\{|f|\le\beta\}} + \operatorname{sgn}(f)\beta\mathbb{1}_{\{|f|\ge\beta\}}$$

and $f_\beta^+ = f - f_\beta$. For every $\sigma = (X_1,\dots,X_k)$ let

$$V_\beta^- = \{(f_\beta^-(X_i))_{i=1}^k : f\in F\},\qquad V_\beta^+ = \{(f_\beta^+(X_i))_{i=1}^k : f\in F\}.$$

Proof of Theorem 6.4. First, by Theorem 6.9, with probability at least $1-\delta-2\exp(-c_1k)$,

$$\mathbb{E}_g \sup_{f\in F\cap c_2/\sqrt{k}B(L_2(\mu))} \left|\sum_{i=1}^{ck} g_i f(X_i)\right| \ge c_3\sqrt{k},$$

where $c > 1$. Set

$$H = F\cap \frac{c_2}{\sqrt{k}}B(L_2(\mu))$$

and note that by the proof of Theorem 4.2 for the class H and $m = ck$, each $h\in H$ can be written as $h = h_1 + h_2$, where $h_1 \in H - H \subset 2H$ (by the convexity and symmetry of F), and $h_2\in H$. Moreover, if we write $H\subset H_1 + H_2$ then with μ^{ck} probability $1-\delta$, $P_\sigma H_1 \subset c_4\gamma_{2,\tau_{ck}}(F, L_2)W_{ck} \subset c_4 AW_{ck}$, where $c_4 = c_4(L,\delta)$. By a standard concentration argument—similar to the one used in Theorem 4.2, since $|H_2| \le \exp(c_5k)$ then for every $h_2\in H_2$, $\|h_2\|_{L_2^{ck}} \le c_6\|h_2\|_{L_2} \le c_7/\sqrt{k}$. Thus, $P_\sigma H_2 \subset (c_7/\sqrt{k})B_2^{ck}$, and since $B_2^{ck}\subset W_{ck}$, then

$$P_\sigma H \subset P_\sigma H_1 + P_\sigma H_2 \subset c_8 W_{ck},$$

where $c_8 = c_8(A, L, \delta)$.

Let $\sigma = (X_i)_{i=1}^{ck}$ for which the above estimates hold, fix β to be named later and let V_β^+ and V_β^- be as in Definition 6.10 for the set H. Consider the set

$$W_k^\beta = \{x\in\mathbb{R}^{ck} : x_i^* \le (c_8/\sqrt{i}) - \beta \text{ for } i \le (c_8/\beta)^2, x_i^* = 0 \text{ for } i > (c_8/\beta)^2\}$$

and observe that $V_\beta^+ \subset W_{ck}^\beta$. Therefore, if we set $i_\beta = (c_8/\beta)^2$ and select β to satisfy that $1 \le i_\beta \le ck$ then,

$$\mathbb{E} \sup_{v \in V_\beta^+} \left| \sum_{i=1}^{ck} g_i v_i \right| \le \mathbb{E} \sup_{w \in W_{ck}^\beta} \left| \sum_{i=1}^{ck} g_i v_i \right| \le \mathbb{E} \sum_{i=1}^{i_\beta} \frac{c_8}{\sqrt{i}} g_i^* \le c_9 \sqrt{i_\beta \log(ek/i_\beta)} \le \frac{c_3}{2}\sqrt{k}$$

for an appropriate choice of $\beta \sim c_3/\sqrt{k}$. Since $V = P_\sigma H \subset V_\beta^+ + V_\beta^-$, then

$$\begin{aligned} c_3\sqrt{k} &\le \mathbb{E}\sup_{v \in V} \left| \sum_{i=1}^{ck} g_i v_i \right| \\ &\le \mathbb{E} \sup_{v \in V_\beta^-} \left| \sum_{i=1}^{ck} g_i v_i \right| + \mathbb{E} \sup_{v \in V_\beta^+} \left| \sum_{i=1}^{ck} g_i v_i \right| \\ &\le \mathbb{E} \sup_{v \in V_\beta^-} \left| \sum_{i=1}^{ck} g_i v_i \right| + \frac{c_3}{2}\sqrt{k}. \end{aligned}$$

Therefore,

$$\mathbb{E} \sup_{v \in \beta^{-1} V_\beta^-} \left| \sum_{i=1}^{ck} g_i v_i \right| \ge \frac{c_3 \sqrt{k}}{2\beta} \ge c_{10} k.$$

Note that

$$\beta^{-1} V_\beta^- = \left\{ \sum_{\{i:\, |f(X_i)| \le \beta\}} \beta^{-1} f(X_i) e_i + \sum_{\{i:\, |f(X_i)| > \beta\}} \mathrm{sgn}(f(X_i)) e_i : f \in H \right\} \subset B_\infty^{ck}.$$

Therefore, by the optimal estimate in the sign-embedding theorem [12], there are constants $c_{11} \sim c_{10}^2$ and $c_{12} \sim c_{10}$ such that

$$\mathrm{VC}(\beta^{-1} V_\beta^-, c_{11}) \ge c_{12} k.$$

In other words, there is a set $I \subset \{1, \ldots, ck\}$, $|I| \ge c_{12} k$ and a vector $(s_i)_{i \in I}$ such that for every $J \subset I$, there is $v_J \in V_\beta^-$ for which

$$\begin{aligned} v_J(i) &\ge s_i + \beta c_{11} \qquad \text{if } i \in J, \\ v_J(i) &\le s_i - \beta c_{11} \qquad \text{if } i \in I \setminus J, \end{aligned}$$

and it is standard to verify that $(s_i)_{i \in I} \subset \beta B_\infty^I$. It remains to show that $(X_i)_{i \in I}$ is $c_{11}\beta$-shattered by F itself. To that end, fix any $J \subset I$, and let $f_J \in F$ be the function for which

$$v_J = \sum_{\{i:\, |f_J(X_i)| \le \beta\}} f_J(X_i) e_i + \sum_{\{i:\, |f_J(X_i)| > \beta\}} \beta \cdot \mathrm{sgn}\,(f_J(X_i)) e_i.$$

Observe that $(I \setminus J) \cap \{i : \mathrm{sgn}(v_J(i)) > 0\} \subset \{i : |f_J(X_i)| \le \beta\}$. Indeed, if there were some $i \in I \setminus J$ for which $\mathrm{sgn}(v_J(i)) > 0$ and $|f_J(X_i)| > \beta$, then on one hand, $v_J(i) = \beta$, but on the other, $v_J(i) \le s_i - \beta c_{11} \le \beta(1 - c_{11}) < \beta$, which is impossible. In a similar fashion, $J \cap \{i : \mathrm{sgn}(v_J(i)) < 0\} \subset \{i : |f_J(X_i)| \le \beta\}$. Finally, fix $i \in J$. If $f_J(X_i) \ne v_J(i)$ and $\mathrm{sgn}(v_J(i)) > 0$ then $f_J(X_i) \ge v_J(i) \ge s_i + \beta c_{11}$. Otherwise, $\mathrm{sgn}(v_J(i)) < 0$, implying that $v_J(i) = f_J(X_i)$. Hence, for every $i \in J$,

$$f_J(X_i) \ge s_i + \beta c_{11},$$

and by the same argument, for every $i \in I \setminus J$,

$$f_J(X_i) \le s_i - \beta c_{11}.$$

Therefore, $\mathrm{VC}(F, \beta c_{11}) = \mathrm{VC}(F, c_{13}/\sqrt{k}) \ge c_{12}k$, as claimed. $\square$

Acknowledgments. The author would like to thank A. Libman, M. Kozdoda and the anonymous referees for their careful reading of the manuscript and for many valuable suggestions and comments.

The research leading to the results presented here has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement [203134], from the Israel Science Foundation Grant 666/06 and from the Australian Research Council Grant DP0986563.

REFERENCES

[1] Alon, N. and Spencer, J. (2000). *The Probabilistic Method*, 2nd ed. Wiley, New York.

[2] de la Peña, V. H. and Giné, E. (1999). *Decoupling: From Dependence to Independence, Randomly Stopped Processes. U-Statistics and Processes. Martingales and Beyond*. Springer, New York. MR1666908

[3] Dudley, R. M. (1999). *Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. MR1720712

[4] Fernique, X. (1975). Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École D'Été de Probabilités de Saint-Flour, IV-1974 Lecture Notes in Math.* **480** 1–96. Springer, Berlin. MR0413238

[5] Gordon, Y., Litvak, A. E., Schütt, C. and Werner, E. (2006). On the minimum of several random variables. *Proc. Amer. Math. Soc.* **134** 3665–3675 (electronic). MR2240681

[6] Gordon, Y., Litvak, A. E., Mendelson, S. and Pajor, A. (2007). Gaussian averages of interpolated bodies and applications to approximate reconstruction. *J. Approx. Theory* **149** 59–73. MR2371614

[7] Haussler, D. (1995). Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik–Chervonenkis dimension. *J. Combin. Theory Ser. A* **69** 217–232. MR1313896

[8] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]* **23**. Springer, Berlin. MR1102015

[9] Litvak, A. E., Milman, V. D., Pajor, A. and Tomczak-Jaegermann, N. (2007). On the Euclidean metric entropy of convex bodies. In *Geometric Aspects of Functional Analysis*. *Lecture Notes in Math.* **1910** 221–235. Springer, Berlin. MR2349608

[10] Matoušek, J. (1995). Tight upper bounds for the discrepancy of half-spaces. *Discrete Comput. Geom.* **13** 593–601. MR1318799

[11] Matoušek, J. (1999). *Geometric Discrepancy: An Illustrated Guide*. *Algorithms and Combinatorics* **18**. Springer, Berlin. MR1697825

[12] Mendelson, S. and Vershynin, R. (2003). Entropy and the combinatorial dimension. *Invent. Math.* **152** 37–55. MR1965359

[13] Mendelson, S., Pajor, A. and Tomczak-Jaegermann, N. (2007). Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.* **17** 1248–1282. MR2373017

[14] Milman, V. D. (1988). Isomorphic symmetrizations and geometric inequalities. In *Geometric Aspects of Functional Analysis (1986/87)*. *Lecture Notes in Math.* **1317** 107–131. Springer, Berlin. MR950978

[15] Milman, V. D. and Schechtman, G. (1986). *Asymptotic Theory of Finite-Dimensional Normed Spaces*. *Lecture Notes in Math.* **1200**. Springer, Berlin. MR856576

[16] Spencer, J. (1985). Six standard deviations suffice. *Trans. Amer. Math. Soc.* **289** 679–706. MR784009

[17] Talagrand, M. (1987). Regularity of Gaussian processes. *Acta Math.* **159** 99–149. MR906527

[18] Talagrand, M. (2005). *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer, Berlin. MR2133757

[19] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

Department of Mathematics
Technion, Israel Institute of Technology
Haifa 32000
Israel
E-mail: shahar.mendelson@anu.edu.au
and
Centre for Mathematics and its Applications
Institute for Advanced Studies
The Australian National University
Canberra, ACT 0200
Australia
E-mail: shahar@tx.technion.ac.il